

Order-sensitivity and equivariance of scoring functions

Tobias Fissler*

*Department of Mathematics
 Imperial College London, United Kingdom
 e-mail: t.fissler@imperial.ac.uk*

and

Johanna F. Ziegel†

*Institute of Mathematical Statistics and Actuarial Science
 University of Bern, Switzerland
 e-mail: johanna.ziegel@stat.unibe.ch*

Abstract: The relative performance of competing point forecasts is usually measured in terms of loss or scoring functions. It is widely accepted that these scoring function should be strictly consistent in the sense that the expected score is minimized by the correctly specified forecast for a certain statistical functional such as the mean, median, or a certain risk measure. Thus, strict consistency opens the way to meaningful forecast comparison, but is also important in regression and M-estimation. Usually strictly consistent scoring functions for an elicitable functional are not unique. To give guidance on the choice of a scoring function, this paper introduces two additional quality criteria. Order-sensitivity opens the possibility to compare two deliberately misspecified forecasts given that the forecasts are ordered in a certain sense. On the other hand, equivariant scoring functions obey similar equivariance properties as the functional at hand – such as translation invariance or positive homogeneity. In our study, we consider scoring functions for popular functionals, putting special emphasis on vector-valued functionals, e.g. the pair (mean, variance) or (Value at Risk, Expected Shortfall).

MSC 2010 subject classifications: Primary 62C99, 62F07; secondary 62G99, 91B06.

Keywords and phrases: Consistency, decision theory, elicibility, equivariance, homogeneity, M-estimation, order-sensitivity, point forecasts, scoring functions, translation invariance.

Received November 2017.

Contents

1	Introduction	1167
2	Analytic properties of expected scores	1172
2.1	Monotonicity	1172

*Tobias Fissler gratefully acknowledges funding of his Chapman Fellowship by the Department of Mathematics at Imperial College London.

†Johanna Ziegel gratefully acknowledges financial support from the Swiss National Science Foundation.

2.2	Self-calibration	1174
3	Order-sensitivity	1176
3.1	Different notions of order-sensitivity	1176
3.2	Componentwise order-sensitivity	1178
3.3	Metrical order-sensitivity	1179
3.3.1	Ratios of expectations with the same denominator	1180
3.3.2	Quantiles	1181
3.3.3	Expectiles	1184
3.4	Order-sensitivity on line segments	1184
3.4.1	The pair (mean, variance)	1185
3.4.2	The pair (Value at Risk, Expected Shortfall)	1187
4	Equivariant functionals and order-preserving scoring functions	1188
4.1	Translation invariance	1190
4.2	Homogeneity	1194
A	Assumptions	1196
B	Auxiliary results	1197
C	Proofs	1198
C.1	Proofs for Section 2	1198
C.2	Proofs for Section 3	1199
C.3	Proofs for Section 4	1206
	Acknowledgments	1209
	References	1209

1. Introduction

From the cradle to the grave, human life is full of decisions. Due to the inherent nature of time, decisions have to be made today, but at the same time, they are supposed to account for unknown and uncertain future events. However, since these future events cannot be *known* today, the best thing to do is to base the decisions on *predictions* for these unknown and uncertain events. The call for and the usage of predictions for future events is literally ubiquitous and even dates back to ancient times. In those days, dreams, divination, and revelation were considered as respected sources for forecasts, with the most prominent example being the Delphic Oracle which was not only consulted for decisions of private life, but also for strategic political decisions concerning peace and war. With the development of natural sciences, mathematics, and in particular statistics and probability theory, the ancient metaphysical art of making qualitative forecasts turned into a sophisticated discipline of science adopting a quantitative perspective. Subfields such as meteorology, mathematical finance, or even futurology evolved.

Acknowledging that forecasts are inherently uncertain, two main questions arise:

- (i) How good is a forecast in *absolute* terms?
- (ii) How good is a forecast in *relative* terms?

While question (i) deals with *forecast validation*, this paper focuses on some aspects of question (ii) which is concerned with *forecast selection*, *forecast comparison*, or *forecast ranking*. Specifically, we present results on order-sensitivity and equivariance of consistent scoring functions for elicitable functionals. These results may provide guidance for choosing a specific scoring function for forecast comparison within the large class of all consistent scoring functions for an elicitable functional of interest.

We adopt the general decision-theoretic framework following Gneiting (2011); cf. Savage (1971); Osband (1985); Lambert, Pennock and Shoham (2008). For some number $n \geq 1$, one has

- (a) observed *ex post* realizations y_1, \dots, y_n of a time series $(Y_t)_{t \in \mathbb{N}}$, taking values in an *observation domain* \mathcal{O} with a σ -algebra \mathcal{O} ;
- (b) a family \mathcal{F} of probability distributions on $(\mathcal{O}, \mathcal{O})$, containing the (conditional) distributions of Y_t ;
- (c) *ex ante* forecasts $x_1^{(i)}, \dots, x_n^{(i)}$, $i \in \{1, \dots, m\}$ of $m \geq 1$ competing experts / forecasters taking values in an *action domain* $\mathcal{A} \subseteq \mathbb{R}^k$ for some $k \geq 1$;
- (d) a *scoring (or loss) function* $S: \mathcal{A} \times \mathcal{O} \rightarrow \mathbb{R}$. The scoring function is assumed to be negatively oriented, that is, if a forecaster reports the quantity $x \in \mathcal{A}$ and $y \in \mathcal{O}$ materializes, she is assigned the *penalty* $S(x, y) \in \mathbb{R}$.

The observations y_t can be real-valued (GDP growth for one year, maximal temperature of one day), vector-valued (wind-speed, weight and height of persons), functional-valued (path of the exchange rate Euro–Swiss franc over one day), or also set-valued (area of rain on a given day, area affected by a flood). In this article, we focus on *point forecasts* that may be vector-valued, which is why we assume $\mathcal{A} \subseteq \mathbb{R}^k$ for some $k \geq 1$ and we equip the Borel set \mathcal{A} with the Borel σ -algebra. One is typically interested in a certain statistical property of the underlying (conditional) distribution F_t of Y_t . We assume that this property can be expressed in terms of a *functional* $T: \mathcal{F} \rightarrow \mathcal{A}$ such as the mean, a certain quantile, or a risk measure. Examples of vector-valued functionals are the covariance matrix of a multivariate observation or a vector of quantiles at different levels. Common examples for scoring functions are the absolute loss $S(x, y) = |x - y|$, the squared loss $S(x, y) = (x - y)^2$ (for $\mathcal{A} = \mathcal{O} = \mathbb{R}$), or the absolute percentage loss $S(x, y) = |(x - y)/y|$ (for $\mathcal{A} = \mathcal{O} = (0, \infty)$).

Forecast comparison is done in terms of *realized scores*

$$\bar{\mathbf{S}}_n^{(i)} = \frac{1}{n} \sum_{t=1}^n S(x_t^{(i)}, y_t), \quad i \in \{1, \dots, m\}. \quad (1.1)$$

That is, a forecaster is deemed to be the better the lower her realized score is. However, there is the following caveat: The forecast ranking in terms of realized scores not only depends on the forecasts and the realizations (as it should definitely be the case), but also on the choice of the scoring function. In order to avoid impure possibilities of manipulating the forecast ranking *ex post* with the data at hand, it is necessary to specify a certain scoring function before the inspection of the data. *A fortiori*, for the sake of transparency and in order

to encourage truthful forecasts, one ought to disclose the choice of the scoring function to the competing forecasters *ex ante*. But still, the optimal choice of the scoring function remains an open problem. One can think of two situations:

- (i) A decision-maker might be aware of their actual economic costs of utilizing misspecified forecasts. In this case, the scoring function should reflect these economic costs.
- (ii) The actual economic costs might be unclear and the scoring function might be just a tool for forecast ranking. However, the directive is given in terms of the functional $T: \mathcal{F} \rightarrow \mathbf{A}$ one is interested in.

For situation (i) described above, one should use the readily economically interpretable cost or scoring function. Therefore, the only concern is situation (ii). In this paper, we consider predictions in a one-period setting, thus, dropping the index t . This is justified by our objectives to understand the properties of scoring functions S which do not change over time and is common in the literature (Murphy and Daan, 1985; Diebold and Mariano, 1995; Lambert, Pennock and Shoham, 2008; Gneiting, 2011).

Assuming the forecasters are *homines oeconomici* and adopting the rationale of expected utility maximization, given a concrete scoring function S , the most sensible action consists in minimizing the expected score $\mathbb{E}_F S(x, Y)$ with respect to the forecast x , where Y follows the distribution F , thus issuing the Bayes act $\arg \min_{x \in \mathbf{A}} \mathbb{E}_F S(x, Y)$. Hence, a scoring function should be incentive compatible in that it encourages truthful and honest forecasts. In line with Murphy and Daan (1985) and Gneiting (2011), we make the following definition.

Definition 1.1 (Consistency and elicibility). A *scoring function* is a map $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ that is \mathcal{F} -integrable.¹ It is \mathcal{F} -consistent for a functional $T: \mathcal{F} \rightarrow \mathbf{A}$ if

$$\bar{S}(T(F), F) \leq \bar{S}(x, F) \quad (1.2)$$

for all $F \in \mathcal{F}$ and for all $x \in \mathbf{A}$, where $\bar{S}(x, F) := \mathbb{E}_F S(x, Y)$. It is *strictly* \mathcal{F} -consistent for T if it is \mathcal{F} -consistent for T and if equality in (1.2) implies $x = T(F)$. A functional $T: \mathcal{F} \rightarrow \mathbf{A}$ is called *elicitable*, if there exists a strictly \mathcal{F} -consistent scoring function for T .

Clearly, elicibility and consistent scoring functions are naturally linked also to estimation problems, in particular, M-estimation (Huber, 1964; Huber and Ronchetti, 2009) and regression with prominent examples being ordinary least squares, quantile, or expectile regression (Koenker, 2005; Newey and Powell, 1987).

The necessity of utilizing strictly consistent scoring functions for meaningful forecast comparison is impressively demonstrated in terms of a simulation study in Gneiting (2011). However, for a given functional $T: \mathcal{F} \rightarrow \mathbf{A}$, there is typically a whole class of strictly consistent scoring functions for it, such as all Bregman functions in case of the mean (Savage, 1971); further examples are given below.

¹We say that a function $a: \mathbf{O} \rightarrow \mathbb{R}$ is \mathcal{F} -integrable if it is F -integrable for each $F \in \mathcal{F}$. A function $g: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ is \mathcal{F} -integrable if $g(x, \cdot)$ is \mathcal{F} -integrable for each $x \in \mathbf{A}$.

Patton (2017) shows that the forecast ranking based on (1.1) may depend on the choice of the strictly consistent scoring function for T in finite samples, and even at the population level if we compare two imperfect forecasts with each other.

Therefore, we naturally have a threefold *elicitation problem*:

- (i) Is T elicitable?
- (ii) What is the class of strictly \mathcal{F} -consistent scoring functions for T ?
- (iii) What are *distinguished* strictly \mathcal{F} -consistent scoring functions for T ?

Even though the denomination and the synopsis of the described problems under the term ‘elicitation problem’ are novel, there is a rich strand of literature in mathematical statistics and economics concerned with the threefold elicitation problem. Foremost, one should mention the pioneering work of Osband (1985), establishing a necessary condition for elicibility in terms of convex level sets of the functional, and a necessary representation of strictly consistent scoring functions, known as *Osband’s principle* (Gneiting, 2011). Whereas the necessity of convex level sets holds in broad generality, Lambert (2013) could specify sufficient conditions for elicibility for functionals taking values in a finite set, and Steinwart et al. (2014) showed sufficiency of convex level sets for real-valued functionals satisfying certain regularity conditions. Moments, ratios of moments, quantiles, and expectiles are in general elicitable, whereas other important functionals such as variance, Expected Shortfall or the mode functional are not (Savage, 1971; Osband, 1985; Weber, 2006; Gneiting, 2011; Heinrich, 2014).

Concerning subproblem (ii) of the elicitation problem, Savage (1971), Reichelstein and Osband (1984), Saaens (2000), and Banerjee, Guo and Wang (2005) gave characterizations for strictly consistent scoring functions for the mean functional of a one-dimensional random variable in terms of Bregman functions. Strictly consistent scoring functions for quantiles have been characterized by Thomson (1979) and Saaens (2000). Gneiting (2011) provides a characterization of the class of strictly consistent scoring functions for expectiles. The case of vector-valued functionals apart from means of random vectors has been treated substantially less than the one-dimensional case (Osband, 1985; Banerjee, Guo and Wang, 2005; Lambert, Pennock and Shoham, 2008; Frongillo and Kash, 2015a,b; Fissler and Ziegel, 2016).

The strict consistency of S only justifies a comparison of two competing forecasts if one of them reports the true functional value. If both of them are misspecified, it is *per se* not possible to draw a conclusion which forecast is ‘closer’ to the true functional value by comparing the realized scores. To this end, some notions of order-sensitivity are desirable. According to Lambert (2013) we say that a scoring function S is \mathcal{F} -order-sensitive for a one-dimensional functional $T: \mathcal{F} \rightarrow \mathbb{A} \subseteq \mathbb{R}$ if for any $F \in \mathcal{F}$ and any $x, z \in \mathbb{A}$ such that either $z \leq x \leq T(F)$ or $z \geq x \geq T(F)$, then $\bar{S}(x, F) \leq \bar{S}(z, F)$. This means, if a forecast lies between the true functional value and some other forecast, then issuing the forecast in-between should yield a smaller expected score than issuing the forecast further away. In particular, order-sensitivity implies consistency. *Vice*

versa, under weak regularity conditions on the functional, strict consistency also implies order-sensitivity if the functional is real-valued; see Nau (1985, Proposition 3), Lambert (2013, Proposition 2), Bellini and Bigozzi (2015, Proposition 3.4).

This article is dedicated to a thorough investigation of order-sensitive scoring functions for vector-valued functionals, thus contributing to a discussion of subproblem (iii) of the elicitation problem. Furthermore, we investigate to which extent invariance or equivariance properties of elicitable functionals are reflected in their respective consistent scoring functions.

Lambert, Pennock and Shoham (2008) introduced a notion of componentwise order-sensitivity for the case of $A \subseteq \mathbb{R}^k$. Friedman (1983) and Nau (1985) considered similar questions in the setting of probabilistic forecasts, coining the term of effectiveness of scoring rules which can be described as order-sensitivity in terms of a metric. In Section 3, we consider three notions of order-sensitivity in the higher-dimensional setting: metrical order-sensitivity, componentwise order-sensitivity, and order-sensitivity on line segments. We discuss their connections (Lemma 3.5) and give conditions when such scoring functions exist (Lemma B.2, Propositions 3.7, 3.8, Corollary 3.16) and of what form they are for the most relevant functionals, such as vectors of quantiles (Propositions 3.11, 3.12, Example 3.14), expectiles (Proposition 3.15), ratios of expectations (Propositions 3.6, 3.9, 3.10, 3.17), the pair of mean and variance (Proposition 3.18, Example 3.19), and the pair consisting of Value at Risk and Expected Shortfall (Proposition 3.20, Example 3.21), two important risk measures in banking and insurance.

Complementing our results on order-sensitivity, in Section 2, we consider the analytic properties of the expected score $x \mapsto \bar{S}(x, F)$, $x \in A \subseteq \mathbb{R}^k$, for some scoring function S and some distribution $F \in \mathcal{F}$. The (strict) consistency of S for some functional T is equivalent the expected score having a (unique) global minimum at $x = T(F)$. Order-sensitivity ensures monotonicity properties of the expected score. As a technical result, we show that under weak regularity assumptions on T , the expected score of a strictly consistent scoring function has a unique local minimum – which, of course, coincides with the global minimum at $x = T(F)$ (Proposition 2.6). Accompanied with a result on self-calibration (Proposition 2.8), a continuity property of the inverse of the expected score, which ensures that the minimum of the expected score is well-separated in the sense of van der Vaart (1998), these two findings may be of interest on their own right in the context of M-estimation (Theorem 2.9).

In Section 4, we consider functionals having an invariance or equivariance property such as translation invariance or homogeneity. It is a natural question whether a functional T that is, for example, translation equivariant has a consistent scoring function that respects this property in the sense that if we evaluate forecast performance of translated predictions and observations, the ranking of predictive performance remains the same as that of the original data. In parametric estimation problems, such a scoring function may allow to translate the data without affecting the estimated parameter values. For one-dimensional functionals, invariance of the scoring function often determines it uniquely up to equivalence while this is not necessarily the case for higher-dimensional func-

tionals (Proposition 4.7 and Corollary 4.12).

In Appendix A, we gather a list of common assumptions, which were originally introduced in Fissler and Ziegel (2016). Appendix B consists of technical results, while all proofs of the main part of this paper are deferred to Appendix C.

2. Analytic properties of expected scores

2.1. Monotonicity

Definition 2.1 (Mixture-continuity). Let \mathcal{F} be convex. A functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ is called *mixture-continuous* if for all $F, G \in \mathcal{F}$ the map

$$[0, 1] \rightarrow \mathbf{A}, \quad \lambda \mapsto T((1 - \lambda)F + \lambda G)$$

is continuous.

It is appealing that one does not have to specify a topology on \mathcal{F} to define mixture-continuity because it suffices to work with the induced Euclidean topology on $[0, 1]$ and on $\mathbf{A} \subseteq \mathbb{R}^k$.

It turns out that mixture-continuity of a functional is strong enough to imply order-sensitivity in the case of one-dimensional functionals (see Nau (1985, Proposition 3), Lambert (2013, Proposition 2), Bellini and Bignozzi (2015, Proposition 3.4)), and desirable monotonicity properties of the expected scores also in higher dimensions (Propositions 2.4 and 2.6). At the same time, numerous functionals of applied relevance are mixture-continuous, and we start by giving examples and a sufficient condition (Proposition 2.2).

It is straight forward to see that the ratio of expectations is mixture-continuous. Moreover, by the implicit function theorem, one can verify the mixture-continuity of quantiles and expectiles directly under appropriate regularity conditions (e.g., in the case of quantiles, all distributions in \mathcal{F} should be C^1 with non-vanishing derivatives). Generalizing Bellini and Bignozzi (2015, Proposition 3.4c), we give a sufficient criterion for mixture-continuity in the next proposition. Our version is not restricted to distributions with compact support (however, the image of the functional must be bounded), and we formulate the result for k -dimensional functionals.

Proposition 2.2. *Let $T: \mathcal{F} \rightarrow \mathbb{R}^k$ be an elicitable functional with a strictly \mathcal{F} -consistent scoring function $S: \mathbb{R}^k \times \mathbf{O} \rightarrow \mathbb{R}$ such that $\bar{S}(\cdot, F)$ is continuous for all $F \in \mathcal{F}$. Then T is mixture-continuous on any $\mathcal{F}_0 \subseteq \mathcal{F}$ such that \mathcal{F}_0 is convex and the image $T(\mathcal{F}_0)$ is bounded.*

Similarly to the original proof of Bellini and Bignozzi (2015), a sufficient criterion for the continuity of $\bar{S}(\cdot, F)$ for any $F \in \mathcal{F}$ is that for all $y \in \mathbf{O}$, the score $S(x, y)$ is quasi-convex and continuous in x .²

²We remark that for $\mathbf{A} \subseteq \mathbb{R}$, if a scoring function S is strictly \mathcal{F}_p -consistent for some functional $T: \mathcal{F}_p \rightarrow \mathbf{A}$ where $\mathcal{F}_p = \{\delta_y: y \in \mathbf{O}\}$ consists of all point measures on \mathbf{O} , then the quasi-convexity of $x \mapsto S(x, y)$ for all $y \in \mathbf{O}$ is equivalent to the \mathcal{F}_p -order-sensitivity of S for T .

Recall that, under appropriate regularity conditions on \mathcal{F} , the asymmetric piecewise linear loss $S_\alpha(x, y) = (\mathbb{1}\{y \leq x\} - \alpha)(x - y)$ and the asymmetric piecewise quadratic loss $S_\tau(x, y) = |\mathbb{1}\{y \leq x\} - \tau|(x - y)^2$ are strictly consistent scoring functions for the α -quantile and the τ -expectile, respectively, and both, S_α as well as S_τ , are continuous in their first argument and convex. Hence, Proposition 2.2 yields that both quantiles and expectiles are mixture-continuous.

Steinwart et al. (2014) used Osband's principle (Osband, 1985) and the assumption of continuity of T with respect to the total variation distance to show order-sensitivity. Bellini and Bignozzi (2015) showed that the weak continuity of a functional T implies its mixture-continuity. Consequently, one can also derive the order-sensitivity in the framework of Steinwart et al. (2014) directly using only mixture-continuity.

Lambert (2013) showed that it is a harder requirement to have order-sensitivity if $T(\mathcal{F})$ is discrete. Then both approaches, invoking Osband's principle or using mixture-continuity, do not work because the interior of the image of T is empty. Moreover, mixture-continuity implies that the functional is constant (such that only trivial cases can be considered). Furthermore, it is proven in Lambert (2013) that for a functional T with a discrete image, all strictly consistent scoring functions are order-sensitive if and only if there is one order-sensitive scoring function for T . In particular, there are functionals admitting strictly consistent scoring functions that are not order-sensitive, one such example being the mode functional.³

Let us turn attention to vector-valued functionals now. To understand the monotonicity properties of the expected score of a mixture-continuous elicitable functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$, it is useful to consider paths $\gamma: [0, 1] \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$, $\gamma(\lambda) = T(\lambda F + (1 - \lambda)G)$ for $F, G \in \mathcal{F}$. If T is elicitable, a classical result asserts that T necessarily has convex level sets (Gneiting, 2011, Theorem 6). This implies that the level sets of γ can only be closed intervals including the case of singletons and the empty set. This rules out loops and some other possible pathologies of γ . Furthermore, under the assumption that T is *identifiable* as defined below, one can even show that the path γ is either injective or constant; see Lemma B.1.

Definition 2.3 (Identifiability). Let $\mathbf{A} \subseteq \mathbb{R}^k$. An \mathcal{F} -integrable function $V: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}^k$ is said to be an \mathcal{F} -*identification function* for a functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ if

$$\bar{V}(T(F), F) = 0$$

for all $F \in \mathcal{F}$. Furthermore, V is a *strict \mathcal{F} -identification function* for T if $\bar{V}(x, F) = 0$ implies $x = T(F)$ for all $F \in \mathcal{F}$ and for all $x \in \mathbf{A}$. A functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ is said to be *identifiable*, if there exists a strict \mathcal{F} -identification

³Note that due to Proposition 1 in Heinrich (2014), the mode functional is elicitable relative to the class of probability measure \mathcal{F} containing unimodal discrete measures. Moreover, interpreting the mode functional as a set-valued functional, it is elicitable in the sense of Gneiting (2011, Definition 2). A strictly \mathcal{F} -consistent scoring function is given by $S(x, y) = \mathbb{1}\{x \neq y\}$. The main result of Heinrich (2014) is that the mode functional is *not* elicitable relative to the class \mathcal{F} of unimodal probability measures with Lebesgue densities.

function for T .

In line with Gneiting (2011, Section 2.4), one can often obtain an identification function as the gradient of a sufficiently smooth scoring function. However, the converse intuition is not so clear – at least in the higher dimensional setting $k > 1$: Not all strict identification functions can be integrated to a strictly consistent scoring function. They have to satisfy the usual integrability conditions (Königsberger, 2004, p. 185); see also Fissler and Ziegel (2016, Corollary 3.3) and the discussion thereafter.

Proposition 2.4. *Let \mathcal{F} be convex and $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ be mixture-continuous and surjective. Let $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ be strictly \mathcal{F} -consistent for T . Then for each $F \in \mathcal{F}$, $t = T(F)$ and each $x \in \mathbf{A}$, $x \neq t$ there is a continuous path $\gamma: [0, 1] \rightarrow \mathbf{A}$ such that $\gamma(0) = x$, $\gamma(1) = t$, and the function $[0, 1] \ni \lambda \mapsto \bar{S}(\gamma(\lambda), F)$ is decreasing. Additionally, for $0 \leq \lambda < \lambda' \leq 1$ such that $\gamma(\lambda) \neq \gamma(\lambda')$ it holds that $\bar{S}(\gamma(\lambda), F) > \bar{S}(\gamma(\lambda'), F)$.*

Remark 2.5. (i) Proposition 2.4 remains valid if S is only \mathcal{F} -consistent. Then, we merely have that the function $[0, 1] \ni \lambda \mapsto \bar{S}(\gamma(\lambda), F)$ is decreasing, so the last inequality in Proposition 2.4 is not necessarily strict.
(ii) If one assumes in Proposition 2.4 that T is also identifiable, one can use the injectivity of γ implied by Lemma B.1 to see that the function $[0, 1] \ni \lambda \mapsto \bar{S}(\gamma(\lambda), F)$ is strictly decreasing.

Under certain (weak) regularity conditions, the expected scores of a strictly consistent scoring function has no other local minimum apart from the global one at $x = T(F)$.

Proposition 2.6. *Let \mathcal{F} be convex and $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ be mixture-continuous and surjective. If $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ is strictly \mathcal{F} -consistent for T , then for all $F \in \mathcal{F}$ the expected score $\bar{S}(\cdot, F): \mathbf{A} \rightarrow \mathbb{R}$ has only one local minimum which is at $x = T(F)$.*

2.2. Self-calibration

With Proposition 2.4 it is possible to prove that, under mild regularity conditions, strictly consistent scoring functions are *self-calibrated* which turns out to be useful in the context of M-estimation.

Definition 2.7 (Self-calibration). A scoring function $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ is called \mathcal{F} -self-calibrated for a functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ with respect to a norm⁴ $\|\cdot\|$ on \mathbf{A} if for all $\varepsilon > 0$ and for all $F \in \mathcal{F}$ there is a $\delta = \delta(\varepsilon, F) > 0$ such that for all $x \in \mathbf{A}$ and $t = T(F)$

$$\bar{S}(x, F) - \bar{S}(t, F) < \delta \quad \implies \quad \|t - x\| < \varepsilon.$$

⁴It is straight forward to use a metric instead of a norm on \mathbf{A} but in this article we only consider $\mathbf{A} \subseteq \mathbb{R}^k$, so we did not see any benefit in considering this more general case. See also the discussion before Definition 3.4.

The notion of self-calibration was introduced by Steinwart (2007) in the context of machine learning. In a preprint version of Steinwart et al. (2014),⁵ the authors translate this concept to the setting of scoring functions as follows (using our notation):

“For self-calibrated S , every δ -approximate minimizer of $\bar{S}(\cdot, F)$, approximates the desired property $T(F)$ with precision not worse than ε . [...] In some sense order sensitivity is a global and qualitative notion while self-calibration is a local and quantitative notion.”

In line with this quotation, self-calibration can be considered as the continuity of the inverse of the expected score $\bar{S}(\cdot, F)$ at the global minimum $x = T(F)$ – and as such, it is a local property of the inverse. This property ensures that convergence of the expected score to its global minimum implies convergence of the forecast to the true functional value. On the other hand, self-calibration of a scoring function S is equivalent to the fact that the argmin $T(F)$ of the expected score $\bar{S}(\cdot, F)$ is a *well-separated* point of minimum in the sense of van der Vaart (1998, p. 45) – as such being a global property of the expected score itself. That means that for any $\varepsilon > 0$

$$\inf\{\bar{S}(x, F) : \|T(F) - x\| \geq \varepsilon\} > \bar{S}(T(F), F).$$

It is relatively straight forward to see that self-calibration implies strict consistency: Let S be \mathcal{F} -self-calibrated for T , $F \in \mathcal{F}$, $t = T(F)$ and $x \in \mathbf{A}$ with $x \neq t$. Then for $\varepsilon := \|x - t\|/2 > 0$ there is a $\delta > 0$ such that $\bar{S}(x, F) - \bar{S}(t, F) \geq \delta > 0$.

In the preprint version of Steinwart et al. (2014) it is shown for $k = 1$ that order-sensitivity implies self-calibration. The next Proposition shows that the kind of order-sensitivity given by Proposition 2.4 also implies self-calibration for $k \geq 1$.

Proposition 2.8. *Let \mathcal{F} be convex, $\mathbf{A} \subseteq \mathbb{R}^k$ be closed, and $T: \mathcal{F} \rightarrow \mathbf{A}$ be a surjective and mixture-continuous functional. If $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ is strictly \mathcal{F} -consistent for T and $\bar{S}(\cdot, F): \mathbf{A} \rightarrow \mathbb{R}$ is continuous for all $F \in \mathcal{F}$, then S is \mathcal{F} -self-calibrated for T .*

We end this subsection about self-calibration by demonstrating its applicability in the context of M-estimation.

Theorem 2.9. *Let $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ be an \mathcal{F} -self-calibrated scoring function for a functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$. Then, the following assertion holds for all $F \in \mathcal{F}$. If Y_1, Y_2, \dots is a sequence of random variables with distribution $F \in \mathcal{F}$ such that*

$$\sup_{x \in \mathbf{A}} \left| \frac{1}{n} \sum_{i=1}^n S(x, Y_i) - \bar{S}(x, F) \right| \xrightarrow{\mathbb{P}} 0,$$

then

$$\arg \min_{x \in \mathbf{A}} \frac{1}{n} \sum_{i=1}^n S(x, Y_i) \xrightarrow{\mathbb{P}} T(F).$$

⁵ Available at <http://users.cecs.anu.edu.au/~williams/papers/P196.pdf>

The proof of Theorem 2.9 is a direct consequence of van der Vaart (1998, Theorem 5.7). Recall that under some additional regularity conditions, it is also possible to derive a Central Limit Theorem associated to the consistency result established in Theorem 2.9. The rate is driven by the dependence structure of the observations Y_1, Y_2, \dots . If they are independent the rate is typically $n^{-1/2}$. The form of the scoring function only enters via the asymptotic covariance. For details, we refer the reader to Chapter 5.3 in van der Vaart (1998). A detailed discussion of the asymptotic covariance and related efficiency considerations of the estimator are beyond the scope of this paper.

3. Order-sensitivity

3.1. Different notions of order-sensitivity

The idea of order-sensitivity is that a forecast lying between the true functional value and some other forecast is also assigned an expected score lying between the two other expected scores. If the action domain is one dimensional, there are only two cases to consider: both forecasts are on the left-hand side of the functional value or on the right-hand side. However, if $A \subseteq \mathbb{R}^k$ for $k \geq 2$, the notion of ‘lying between’ is ambiguous. Two obvious interpretations for the multidimensional case are the componentwise interpretation and the interpretation that one forecast is the convex combination of the true functional value and the other forecast.

Definition 3.1 (Componentwise order-sensitivity). A scoring function $S: A \times O \rightarrow \mathbb{R}$ is called componentwise \mathcal{F} -order-sensitive for a functional $T: \mathcal{F} \rightarrow A \subseteq \mathbb{R}^k$, if for all $F \in \mathcal{F}$, $t = T(F)$ and for all $x, z \in A$ we have that:

$$\begin{aligned} \text{For all } m \in \{1, \dots, d\} : z_m \leq x_m \leq T_m(F) \text{ or } z_m \geq x_m \geq T_m(F) \\ \implies \bar{S}(x, F) \leq \bar{S}(z, F). \end{aligned} \quad (3.1)$$

Moreover, S is called strictly componentwise \mathcal{F} -order-sensitive for T if S is componentwise \mathcal{F} -order-sensitive and if $x \neq z$ in (3.1) implies that $\bar{S}(x, F) < \bar{S}(z, F)$.

Remark 3.2. In economic terms, a strictly componentwise order-sensitive scoring function rewards *Pareto improvements*⁶ in the sense that improving the prediction performance in one component without deteriorating the prediction ability in the other components results in a lower expected score.

⁶The definition of the *Pareto principle* according to Scott and Marshall (2009): “A principle of welfare economics derived from the writings of Vilfredo Pareto, which states that a legitimate welfare improvement occurs when a particular change makes at least one person better off, without making any other person worse off. A market exchange which affects nobody adversely is considered to be a ‘Pareto-improvement’ since it leaves one or more persons better off. ‘Pareto optimality’ is said to exist when the distribution of economic welfare cannot be improved for one individual without reducing that of another.”

Definition 3.3 (Order-sensitivity on line segments). Let $\|\cdot\|$ be the Euclidean norm on \mathbb{R}^k . A scoring function $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ is \mathcal{F} -order-sensitive on line segments for a functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$, if for all $F \in \mathcal{F}$, $t = T(F)$, and for all $v \in \mathbb{S}^{k-1} := \{x \in \mathbb{R}^k: \|x\| = 1\}$ the map

$$\psi: D = \{s \in [0, \infty): t + sv \in \mathbf{A}\} \rightarrow \mathbb{R}, \quad s \mapsto \bar{S}(t + sv, F)$$

is increasing. If the map ψ is strictly increasing, we call S strictly \mathcal{F} -order-sensitive on line segments for T .

These two notions of order-sensitivity do not allow for a comparison of any two misspecified forecasts, no matter where they are relative to the true functional value. An intuitive requirement could be ‘the closer to the true functional value the smaller the expected score’, thus calling for the notion of a metric. Since, for a fixed functional T and some fixed distribution F , we always have a fixed reference point $T(F)$ and we have the induced vector-space structure of \mathbb{R}^k on \mathbf{A} , we shall only work with ℓ^p -norms $\|\cdot\|_p$, $p \in [1, \infty]$. Recall that for $x \in \mathbb{R}^k$, $\|x\|_p := (\sum_{i=1}^k |x_i|^p)^{1/p}$ for $p \in [1, \infty)$ and $\|x\|_\infty := \sup_{i=1, \dots, k} |x_i|$. If the assertion does not depend on the choice of p , we shall usually omit the p in the notation. For other choices of \mathbf{A} , it would be also interesting to replace the norm by a metric in the following definition.

Definition 3.4 (Metrical order-sensitivity). Let $p \in [1, \infty]$. A scoring function $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ is *metrically \mathcal{F} -order-sensitive* for a functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ relative to the ℓ^p -norm, if for all $F \in \mathcal{F}$, $t = T(F)$ and for all $x, z \in \mathbf{A}$ we have that

$$\|x - t\|_p \leq \|z - t\|_p \implies \bar{S}(x, F) \leq \bar{S}(z, F). \quad (3.2)$$

If additionally the inequalities in (3.2) are strict, we say that S is *strictly metrically \mathcal{F} -order-sensitive* for T relative to $\|\cdot\|_p$.

Similarly to (strict) consistency, all three notions of (strict) order-sensitivity are preserved when considering two scoring functions that are equivalent.⁷

The notion of componentwise order-sensitivity corresponds almost literally to the notion of accuracy-rewarding scoring functions introduced by Lambert, Pennock and Shoham (2008). Metrically order-sensitivity scoring functions have their counterparts in the field of probabilistic forecasting in effective scoring rules introduced by Friedman (1983) and further investigated by Nau (1985). Actually, the latter paper has also given the inspiration for the notion of order-sensitivity on line segments. It is obvious that any of the three notions of (strict) order-sensitivity implies (strict) consistency. The next lemma formally states this result and gives some logical implications concerning the different notions of order-sensitivity. The proof is standard and therefore omitted.

⁷Two scoring functions $S_1, S_2: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ are *equivalent* if there is a positive constant $\lambda > 0$ and an \mathcal{F} -integrable function $a: \mathbf{O} \rightarrow \mathbb{R}$ such that $S_2(x, y) = \lambda S_1(x, y) + a(y)$, for all $(x, y) \in \mathbf{A} \times \mathbf{O}$.

Lemma 3.5. *Let $T: \mathcal{F} \rightarrow A \subseteq \mathbb{R}^k$ be a functional and $S: A \times \mathcal{O} \rightarrow \mathbb{R}$ a scoring function.*

- (i) *Let $p \in [1, \infty)$. If S is (strictly) metrically \mathcal{F} -order-sensitive for T relative to the ℓ^p -norm, then S is (strictly) componentwise \mathcal{F} -order-sensitive for T .*
- (i') *If S is (strictly) metrically \mathcal{F} -order-sensitive for T relative to the ℓ^∞ -norm, then S is componentwise \mathcal{F} -order-sensitive for T .*
- (i'') *If S is (strictly) metrically \mathcal{F} -order-sensitive for T relative to the ℓ^∞ -norm, then S is (strictly) \mathcal{F} -consistent for T .*
- (ii) *If S is (strictly) componentwise \mathcal{F} -order-sensitive for T , then S is (strictly) \mathcal{F} -order-sensitive on line segments for T .*
- (iii) *If S is (strictly) \mathcal{F} -order-sensitive on line segments for T , then S is (strictly) \mathcal{F} -consistent for T .*

3.2. Componentwise order-sensitivity

Under restrictive regularity assumptions, Lambert, Pennock and Shoham (2008, Theorem 5) claim that whenever a functional has a componentwise order-sensitive scoring function, the components of the functional must be elicitable. Moreover, assuming that the measures in \mathcal{F} have finite support, they assert that any componentwise order-sensitive scoring function is the sum of strictly consistent scoring functions for the components. Lemma B.2 shows the first claim under less restrictive smoothness assumptions on the scoring function. For many common examples of functionals, the second claim can be shown relaxing the restrictive condition on \mathcal{F} .

If $T_m: \mathcal{F} \rightarrow A_m \subseteq \mathbb{R}$, $m \in \{1, \dots, k\}$, are mixture-continuous and elicitable with strictly \mathcal{F} -consistent scoring functions $S_m: A_m \times \mathcal{O} \rightarrow \mathbb{R}$, then they are order-sensitive according to Lambert (2013, Proposition 2) and Bellini and Big-nozzi (2015, Proposition 3.4). Therefore, the sum $\sum_{m=1}^k S_m(x_m, y)$ is strictly componentwise \mathcal{F} -order-sensitive for (T_1, \dots, T_k) . More interestingly, one can establish the reverse of the last assertion. Any strictly componentwise order-sensitive scoring function must necessarily be additively separable. In Fissler and Ziegel (2016, Section 4), we established a dichotomy for functionals with elicitable components: In most relevant cases, the functional (the corresponding strict identification function, respectively) satisfies Assumption (V4) therein (e.g., when the functional is a vector of different quantiles and / or different expectiles with the exception of the 1/2-expectile), or it is a vector of ratios of expectations with the same denominator, or it is a combination of both situations. Under some regularity conditions, Fissler and Ziegel (2016, Propositions 4.2 and 4.4) characterize the form of strictly consistent scoring functions for the first two situations, whereas Fissler and Ziegel (2016, Remark 4.5) is concerned with the third situation. For this latter situation, any strictly consistent scoring function must be necessarily additive for the respective blocks of the functional. And for the first situation, Fissler and Ziegel (2016, Proposition 4.2) yields the additive form of S automatically. It remains to consider the case of Fissler and

Ziegel (2016, Proposition 4.4), that is, a vector of ratios of expectations with the same denominator.

Proposition 3.6. *Let $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ be a ratio of expectations with the same denominator, that is, $T(F) = \mathbb{E}_F[p(Y)]/\mathbb{E}_F[q(Y)]$ for some \mathcal{F} -integrable functions $p: \mathbf{O} \rightarrow \mathbb{R}^k$, $q: \mathbf{O} \rightarrow \mathbb{R}$ such that $\mathbb{E}_F[q(Y)] > 0$ for all $F \in \mathcal{F}$.⁸ Assume that T is surjective, and that $\text{int}(\mathbf{A}) \neq \emptyset$ is simply connected. Moreover, consider the strict \mathcal{F} -identification function $V: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}^k$, $V(x, y) = q(y)x - p(y)$ and some strictly \mathcal{F} -consistent scoring function $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ such that the Assumptions (V1), (S2), (F1), and (VS1) hold. If S is strictly componentwise \mathcal{F} -order-sensitive for T , then S is of the form*

$$S(x_1, \dots, x_k, y) = \sum_{m=1}^k S_m(x_m, y), \quad (3.3)$$

for almost all $(x, y) \in \mathbf{A} \times \mathbf{O}$, where $S_m: \mathbf{A}_m \times \mathbf{O} \rightarrow \mathbb{R}$, $m \in \{1, \dots, k\}$, are strictly \mathcal{F} -consistent scoring functions for $T_m: \mathcal{F} \rightarrow \mathbf{A}_m$, $\mathbf{A}_m := T_m(\mathcal{F}) \subseteq \mathbb{R}$, and $T_m(F) = \mathbb{E}_F[p_m(Y)]/\mathbb{E}_F[q(Y)]$.

The notion of componentwise order-sensitivity has an appealing interpretation in the sense that it rewards Pareto improvements of the predictions; see Remark 3.2. The results of Lemma B.2 and Proposition 3.6 give a clear understanding of the concept including its limitations to the case of functionals only consisting of elicitable components.

Ehm et al. (2016) introduced Murphy diagrams for forecast comparison of quantiles and expectiles. Murphy diagrams have the advantage that forecasts are compared simultaneously with respect to all consistent scoring functions for the respective functional. For many multivariate functionals such as ratios of expectations, the methodology cannot be readily extended because there are no mixture representations available for the class of all consistent scoring functions. Proposition 3.6 shows that when considering only componentwise order-sensitive consistent scoring functions, the situation is different and mixture representations (and hence Murphy diagrams) are readily available for forecast comparison.

3.3. Metrical order-sensitivity

For a real-valued functional T there can be at most one strictly metrically order-sensitive scoring function, up to equivalence. To show this, we use Osband's principle and impose the corresponding regularity conditions.

⁸It is no loss of generality to assume that $\bar{q}(F) > 0$ for all $F \in \mathcal{F}$ in Proposition 3.6. In order to ensure that T is well-defined, necessarily $\bar{q}(F) \neq 0$ for all $F \in \mathcal{F}$. However, Assumption (V1) implies that \mathcal{F} is convex. So if there are $F_1, F_2 \in \mathcal{F}$ such that $\bar{q}(F_1) < 0$ and $\bar{q}(F_2) > 0$ then there is a convex combination G of F_1 and F_2 such that $\bar{q}(G) = 0$. Consequently, either $\bar{q}(F) > 0$ for all $F \in \mathcal{F}$ or $\bar{q}(F) < 0$ for all $F \in \mathcal{F}$, and by possibly changing the sign of p one can assume that the first case holds.

Proposition 3.7. *Let $T: \mathcal{F} \rightarrow A \subseteq \mathbb{R}$ be a surjective, elicitable and identifiable functional with an oriented strict \mathcal{F} -identification function $V: A \times \mathcal{O} \rightarrow \mathbb{R}$. If $\text{int}(A) \neq \emptyset$ is convex and $S, S^*: A \times \mathcal{O} \rightarrow \mathbb{R}$ are two strictly metrically \mathcal{F} -order-sensitive scoring functions for T such that the Assumptions (V1), (V2), (S1), (F1) and (VS1) (with respect to both scoring functions) hold, then S and S^* are equivalent almost everywhere.*

For the higher-dimensional setting we can show a slightly more limited version of Proposition 3.7. Two scoring functions that are additively separable as in (3.3) and that are strictly metrically order-sensitive for the same functional must necessarily be equivalent. For most practically relevant cases – namely when we consider an ℓ^p -norm with $p \in [1, \infty)$ and when the functional possesses an identification function satisfying Assumption (V4) or that are ratios of expectations with the same denominator – Lemma 3.5, Proposition 3.6 and Fissler and Ziegel (2016, Proposition 4.2) yield that any metrically order-sensitive scoring function – presuming there is one – is additively separable. Hence, for these situations, metrically order-sensitive scoring functions are unique, up to equivalence.

Proposition 3.8. *Let $S: A \times \mathcal{O} \rightarrow \mathbb{R}$ be a strictly metrically \mathcal{F} -order-sensitive scoring function for a surjective functional $T = (T_1, \dots, T_k): \mathcal{F} \rightarrow A \subseteq \mathbb{R}^k$ of the form*

$$S(x_1, \dots, x_k, y) = \sum_{m=1}^k S_m(x_m, y)$$

for all $(x, y) \in A \times \mathcal{O}$ where $S_m: A_m \times \mathcal{O} \rightarrow \mathbb{R}$, $m \in \{1, \dots, k\}$, $A_m = \{x_m \in \mathbb{R}: \exists(z_1, \dots, z_k) \in A \text{ and } z_m = x_m\}$, are strictly \mathcal{F} -consistent scoring functions for T_m . Assume that $\text{int}(A) \neq \emptyset$. Then, the following assertions hold:

- (i) *The scoring functions S_m , $m \in \{1, \dots, k\}$, are strictly metrically \mathcal{F} -order-sensitive for T_m .*
- (ii) *Let $\lambda_1, \dots, \lambda_k > 0$ and define the scoring function $S^*: A \times \mathcal{O} \rightarrow \mathbb{R}$ via*

$$S^*(x_1, \dots, x_k) = \sum_{m=1}^k \lambda_m S_m(x_m, y).$$

Then S^ is strictly metrically \mathcal{F} -order-sensitive (with respect to the same ℓ^p -norm as S) if and only if $\lambda_1 = \dots = \lambda_k$.*

Next, we use the derived theoretical results to examine when some popular functionals admit strictly metrically order-sensitive scoring functions, and if so, of what form they are.

3.3.1. Ratios of expectations with the same denominator

We start with the one-dimensional characterization.

Proposition 3.9. *Let \mathcal{F} be convex and $p, q: \mathcal{O} \rightarrow \mathbb{R}$ two \mathcal{F} -integrable functions such that $\bar{q}(F) > 0$ for all $F \in \mathcal{F}$. Define $T: \mathcal{F} \rightarrow A \subseteq \mathbb{R}$, $T(F) = \bar{p}(F)/\bar{q}(F)$*

and assume that T is surjective and $\text{int}(\mathbf{A}) \neq \emptyset$ is convex. Then the following two assertions are true:

- (i) Any scoring function which is equivalent to

$$S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}, \quad S(x, y) = \frac{1}{2}q(y)x^2 - p(y)x \quad (3.4)$$

is strictly metrically \mathcal{F} -order-sensitive for T .

- (ii) If \mathcal{F} is such that Assumptions (V1), (F1) are satisfied with $V(x, y) = q(y)x - p(y)$, then any scoring function $S^*: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$, which is strictly metrically \mathcal{F} -order-sensitive and satisfies Assumptions (S1) and (VS1), is equivalent to S defined at (3.4) almost everywhere.

Now, we turn to the multivariate characterization.

Proposition 3.10. Let $k \geq 2$, \mathcal{F} be convex and $p: \mathbf{O} \rightarrow \mathbb{R}^k$, $q: \mathbf{O} \rightarrow \mathbb{R}$ two \mathcal{F} -integrable functions such that $\bar{q}(F) > 0$ for all $F \in \mathcal{F}$. Define $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$, $T(F) = \bar{p}(F)/\bar{q}(F)$ and assume that T is surjective and $\text{int}(\mathbf{A}) \neq \emptyset$. Then, the following assertions are true:

- (i) Any scoring function which is equivalent to

$$S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}, \quad S(x_1, \dots, x_k, y) = \sum_{m=1}^k \frac{1}{2}q(y)x_m^2 - p_m(y)x_m \quad (3.5)$$

is strictly metrically \mathcal{F} -order-sensitive for T with respect to the ℓ^2 -norm.

- (ii) If \mathcal{F} is such that Assumptions (V1), (F1) are satisfied with $V(x, y) = q(y)x - p(y)$, then any scoring function $S^*: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$, which is strictly metrically \mathcal{F} -order-sensitive with respect to the ℓ^2 -norm and satisfies Assumptions (S1) and (VS1), is equivalent to S defined at (3.5) almost everywhere.
- (iii) If \mathcal{F} is such that Assumptions (V1), (F1) are satisfied with $V(x, y) = q(y)x - p(y)$, then there is no scoring function $S^*: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ which satisfies Assumptions (S1) and (VS1) and which is strictly metrically \mathcal{F} -order-sensitive with respect to an ℓ^p -norm with $p \in [1, \infty) \setminus \{2\}$.

Savage (1971, Section 5) has already shown that in case of the mean, the squared loss is essentially the only symmetric loss in the sense that it is the only metrically order-sensitive loss for the mean. See also Patton (2017, Section 2.1) for a discussion that symmetry – or metrical order-sensitivity – is not necessary for strict consistency of scoring functions with respect to the mean.

3.3.2. Quantiles

Since we treat only point-valued functionals in this article, we shall assume that the α -quantile of F is a singleton and identify the set with its unique element

(henceforth, we shall refer to this assumption as F having a unique α -quantile).⁹ Furthermore, note that assuming the identifiability of the α -quantile with the canonical identification function $V_\alpha(x, y) = \mathbb{1}\{y \leq x\} - \alpha$ on a class \mathcal{F} amounts to assuming that $F(q_\alpha(F)) = \alpha$ for all $F \in \mathcal{F}$.¹⁰

Proposition 3.11. *Let $\alpha \in (0, 1)$ and \mathcal{F} be a family of distribution functions F on \mathbb{R} with unique α -quantiles $q_\alpha(F)$ satisfying $F(q_\alpha(F)) = \alpha$ for all $F \in \mathcal{F}$. Assume that there is an $F_0 \in \mathcal{F}$, such that its translation $F_\lambda(\cdot) = F_0(\cdot - \lambda)$ is also an element of \mathcal{F} for all $\lambda \in \mathbb{R}$. Consequently, $T_\alpha = q_\alpha: \mathcal{F} \rightarrow \mathbb{A} = \mathbb{R}$ is surjective. Under assumptions (V1) with respect to the strict identification function $V_\alpha: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $V_\alpha(x, y) = \mathbb{1}\{y \leq x\} - \alpha$, there is no strictly metrically \mathcal{F} -order-sensitive scoring function for T_α satisfying Assumption (S1).*

The reasons for the non-existence of a strictly metrically order-sensitive scoring function for the α -quantile are of different nature in the two cases that $\alpha \neq 1/2$ and that $\alpha = 1/2$ in the proof of Proposition 3.11. In both cases, we used Osband's principle to derive a representations of the derivative of the expected score. Assuming that the derivative has the form as stated in Osband's principle, one can directly derive a contradiction for $\alpha \neq 1/2$. However, for $\alpha = 1/2$, this form merely implies that the distributions in \mathcal{F} must be symmetric around their medians. This is not contradictory to the form of the gradient derived via Osband's principle, but only to the assumption that \mathcal{F} is convex. Dropping this assumption, we can derive the following Lemma. The proof is straight forward from Lemma B.3.

Proposition 3.12. *Let \mathcal{F} be a family of distribution functions on \mathbb{R} with unique medians $T_{1/2}: \mathcal{F} \rightarrow \mathbb{R}$ and finite first moments. If all distributions in \mathcal{F} are symmetric around their medians in the sense that*

$$F(T_{1/2}(F) + x) = 1 - F((T_{1/2}(F) - x) -) \quad (3.6)$$

for all $F \in \mathcal{F}$, $x \in \mathbb{R}$, then any scoring function that is equivalent to the absolute loss $S: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $S(x, y) = |x - y|$, is strictly metrically \mathcal{F} -order-sensitive with respect to the median.

⁹Recall that the α -quantile of a distribution F consists of all points $x \in \mathbb{R}$ satisfying $\lim_{t \uparrow x} F(t) \leq \alpha \leq F(x)$.

¹⁰Actually, assuming \mathcal{F} is convex and rich enough, this holds for any identification function for the α -quantile. Indeed, consider some distribution function $F_0 \in \mathcal{F}$ and some level $\alpha \in (0, 1)$. Fix some $x_0 \in \mathbb{R}$ such that $F_0(x_0) < \alpha$, implying that $q_\alpha(F_0) > x_0$. Assume that for any $\lambda \in [0, 1]$, the distribution

$$F_\lambda(x) = \begin{cases} F_0(x), & x < x_0 \\ (1 - \lambda)F_0(x) + \lambda, & x \geq x_0 \end{cases}$$

is an element of \mathcal{F} . Then, there is some $\lambda' \in (0, 1)$ such that $F_{\lambda'}(x_0) = \alpha$ implying that $F_\lambda(x_0) > \alpha$ for all $\lambda \in (\lambda', 1]$ and $q_\alpha(F_\lambda) = x_0$ for all $\lambda \in [\lambda', 1]$. Assume that V is a strict \mathcal{F} -identification function for q_α . That means $\bar{V}(x_0, F_\lambda) = 0$ for all $\lambda \in [\lambda', 1]$ and $\bar{V}(x_0, F_\lambda) \neq 0$ for all $\lambda \in [0, \lambda')$. Consider some $\lambda \in [\lambda', 1]$. Then,

$$\bar{V}(x_0, F_\lambda) = (1 - \lambda)\bar{V}(x_0, F_0) + \lambda\bar{V}(x_0, F_1) = (1 - \lambda)\bar{V}(x_0, F_0) \neq 0.$$

This is a contradiction to V being a strict \mathcal{F} -identification function for q_α .

As mentioned above, under the conditions of Proposition 3.12, the necessary characterization of strictly consistent scoring functions via Osband's principle is not available. In particular, this means that we cannot use Proposition 3.7. Indeed, if the distributions in \mathcal{F} are symmetric around their medians in the sense of (3.6) and under the integrability condition that all elements in \mathcal{F} have a finite first moment, the median and the mean coincide. Hence, any convex combination of a strictly consistent scoring function for the mean and the median provides a strictly consistent scoring function. *A fortiori*, any scoring function which is equivalent to $S(x, y) = (1 - \lambda)|x - y| + \lambda|x - y|^2$, $\lambda \in [0, 1]$ is strictly metrically \mathcal{F} -order-sensitive. However, the class of strictly metrically \mathcal{F} -order-sensitive scoring functions is even bigger – Lehmann and Casella (1998, Corollary 7.19, p. 50) show that (subject to integrability conditions) for an even and strictly convex function $\Phi: \mathbb{R} \rightarrow \mathbb{R}$, the score $S(x, y) = \Phi(x - y)$ is strictly metrically \mathcal{F} -order-sensitive for the median. Note that if the distributions in \mathcal{F} are symmetric, their *center of symmetry*, which is the functional solving (3.6), is unique (Fissler, 2017, Lemma 4.1.34), even if the median is not unique. The result of Lehmann and Casella (1998, Corollary 7.19, p. 50) holds for this center of symmetry. Acknowledging that some popular choices for Φ are not strictly convex (see Example 3.14), the following proposition gives a refinement of their result.

Proposition 3.13. *Let \mathcal{F} be a class of symmetric distributions on \mathbb{R} with center of symmetry $C: \mathcal{F} \rightarrow \mathbb{R}$, that is, $F(C(F) + x) = 1 - F((C(F) - x)-)$ for all $F \in \mathcal{F}$, $x \in \mathbb{R}$. Let $\Phi: \mathbb{R} \rightarrow \mathbb{R}$ be a convex and even function, and $S: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $S(x, y) = \Phi(x - y)$. For any $x \in \mathbb{R}$, define the function $\Psi_x: \mathbb{R} \rightarrow \mathbb{R}$, $\Psi_x(y) = \frac{1}{2}(\Phi(x - y) + \Phi(-x - y))$, and for $x, z \in \mathbb{R}$ the set $M_{x,z} = \{y \in \mathbb{R}: \Psi_x(y) - \Psi_z(y) > 0\}$. If for all $F \in \mathcal{F}$ and for all $x, z \in \mathbb{R}$ with $|x| > |z|$ one has that $\mathbb{P}(Y - C(F) \in M_{x,z}) > 0$, $Y \sim F$, then S is strictly metrically \mathcal{F} -order-sensitive for C . In particular, if for all $F \in \mathcal{F}$ and for all $x \neq 0$ it holds that $\mathbb{P}(Y - C(F) \in M_{x,0}) > 0$, $Y \sim F$, then S is strictly \mathcal{F} -consistent for C .*

If Φ is strictly convex then $M_{x,z} = \mathbb{R}$ for all $|x| > |z|$.

Example 3.14. Let \mathcal{F} be a class of symmetric distributions and $S(x, y) = \Phi(x - y)$.

- (i) If $\Phi(t) = |t|^2$, the squared loss arises. Since Φ is strictly convex, the squared loss is strictly metrically \mathcal{F} -order-sensitive.
- (ii) For $\Phi(t) = |t|$, S takes the form of the absolute loss. Then S is strictly metrically \mathcal{F} -order-sensitive (and strictly \mathcal{F} -consistent) if and only if $C(F) \in \text{supp}(F)$ for all $F \in \mathcal{F}$.¹¹
- (iii) Another prominent example of a metrically order-sensitive scoring function for the center of a symmetric distribution besides the absolute or the squared loss is the so-called *Huber loss* which was presented in Huber

¹¹With the support of F $\text{supp}(F)$ we denote the support of the measure induced by F . In this context, $C(F) \in \text{supp}(F)$ is equivalent to F having a unique median.

(1964) and arises upon taking $S(x, y) = \Phi(x - y)$ with

$$\Phi(t) = \begin{cases} \frac{1}{2}t^2, & \text{for } |t| < k, \\ k|t| - \frac{1}{2}k^2, & \text{for } |t| \geq k, \end{cases}$$

where $k \in \mathbb{R}$, $k \geq 0$ is a tuning parameter. The Huber loss is strictly metrically \mathcal{F} -order-sensitive (strictly \mathcal{F} -consistent) if and only if $[C(F) - k, C(F) + k] \cap \text{supp}(F) \neq \emptyset$ for all $F \in \mathcal{F}$.

We emphasize that there are not only metrically-order sensitive strictly consistent scoring functions for the center of symmetric distributions. One can also use asymmetric scoring functions, for example those for the median or the mean, to elicit the center of symmetry.

Due to the negative result of Proposition 3.11 we dispense with an investigation of scoring functions that are metrically order-sensitive for vectors of different quantiles.

3.3.3. Expectiles

The special situation of the $1/2$ -expectile, which coincides with the mean functional, was already considered in Subsection 3.3.1, so let $\tau \neq 1/2$. It is obvious that the canonical scoring function for the τ -expectile, that is, the asymmetric squared loss

$$S_\tau(x, y) = |\mathbb{1}\{y \leq x\} - \tau|(x - y)^2$$

is not metrically order-sensitive since $x \mapsto S_\tau(x + y, y)$ is not an even function. *A fortiori*, it turns out that (under some assumptions) there is no strictly metrically \mathcal{F} -order-sensitive scoring function for the τ -expectile for $\tau \neq 1/2$.

Proposition 3.15. *Let $\tau \in (0, 1)$, $\tau \neq 1/2$, and $T_\tau = \mu_\tau: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}$, $\text{int}(\mathbf{A}) \neq \emptyset$ convex, be the τ -expectile. Assume that T_τ is surjective, and that Assumption (V1) holds with respect to the strict \mathcal{F} -identification function $V_\tau(x, y) = 2|\mathbb{1}\{y \leq x\} - \tau|(x - y)$. Suppose that $\bar{V}(\cdot, F)$ is twice differentiable for all $F \in \mathcal{F}$ and that there is a strictly \mathcal{F} -consistent scoring function $S: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$ such that $\bar{S}(\cdot, F)$ is three times differentiable for all $F \in \mathcal{F}$. In particular, let each $F \in \mathcal{F}$ be differentiable with derivative $f = F'$.*

If there is a $t \in \mathbf{A}$ and $F_1, F_2 \in \mathcal{F}$ such that $T_\tau(F_1) = T_\tau(F_2) = t$, $F_1(t) = F_2(t)$, but $F_1'(t) = f_1(t) \neq f_2(t) = F_2'(t)$, then S is not metrically \mathcal{F} -order-sensitive.

Interestingly, the arguments provided in the proof of Proposition 3.15 leads to an alternative proof that the squared loss is the only strictly metrically order-sensitive scoring function for the mean, up to equivalence; see Remark C.1 for details.

3.4. Order-sensitivity on line segments

Recalling Lemma 3.5, every componentwise order-sensitive scoring function is also order-sensitive on line segments. However, for the particular class of *linear*

functionals, the following corollary shows that any strictly consistent scoring function is already strictly componentwise order-sensitive on line segments.¹²

Corollary 3.16. *If \mathcal{F} is convex and $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ is linear and surjective, then any strictly \mathcal{F} -consistent scoring function for T is strictly \mathcal{F} -order-sensitive on line segments.*

Corollary 3.16 immediately leads the way to the result that the class of strictly order-sensitive scoring functions on line segments is strictly bigger than the class of strict componentwise order-sensitive scoring functions (for some functionals with dimension $k \geq 2$.) E.g. consider a vector of expectations satisfying the conditions of Proposition 3.6 which are the same as the ones in Fissler and Ziegel (2016, Proposition 4.4). Due to the latter result, there are strictly consistent scoring functions – and hence, with Corollary 3.16, strictly order-sensitive on line segments – which are not additively separable. By Proposition 3.6 they cannot be strictly componentwise order-sensitive.

We can extend the result of Corollary 3.16 to the case of ratios of expectations with the same denominator.

Proposition 3.17. *Let $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ be a ratio of expectations with the same denominator, that is, $T(F) = \bar{p}(F)/\bar{q}(F)$ for some \mathcal{F} -integrable functions $p: \mathbf{O} \rightarrow \mathbb{R}^k$, and $q: \mathbf{O} \rightarrow \mathbb{R}$ where we assume that $\bar{q}(F) > 0$ for all $F \in \mathcal{F}$ and that \mathbf{A} is open and convex. Any scoring function of the form*

$$S(x, y) = -\phi(x)q(y) + \nabla\phi(x)(q(y)x - p(y)) \quad (3.7)$$

is strictly \mathcal{F} -order sensitive on line segments, where ϕ is strictly convex differentiable function on \mathbf{A} .

Fissler and Ziegel (2016, Proposition 4.4) shows that essentially all strictly consistent scoring functions for T in the above Proposition 3.17 are of the form at (3.7); see also Frongillo and Kash (2015a, Theorem 13).

Order-sensitivity on line segments is stable under applying an isomorphism via the revelation principle (Gneiting, 2011, Theorem 4). However, dropping the linearity assumption on the bijection in the revelation principle, order-sensitivity on line segments is generally not preserved; see Subsection 3.4.1.

3.4.1. The pair (mean, variance)

The pair (mean, variance) is of importance not only from an applied point of view but it is also an interesting example in the theory about elicibility. Due to the lack of convex level sets, variance is not elicitable (Gneiting, 2011, Theorem 6). However, the pair (mean, variance) is a bijection of the (elicitable) pair

¹²According to Abernethy and Frongillo (2012), we call a functional $T: \mathcal{F} \rightarrow \mathbf{A}$ linear, if it behaves linearly for mixtures of distributions. That is, for any $F, G \in \mathcal{F}$ such that $(1 - \lambda)F + \lambda G \in \mathcal{F}$ for $\lambda \in [0, 1]$ it holds that $T((1 - \lambda)F + \lambda G) = (1 - \lambda)T(F) + \lambda T(G)$. Examples of linear functionals are expectations of transformations, that is, $T(F) = \mathbb{E}_F[p(Y)]$ for some \mathcal{F} -integrable function $p: \mathbf{O} \rightarrow \mathbb{R}^k$.

(mean, second moment), and, invoking the revelation principle (Gneiting, 2011, Theorem 4), variance is *jointly* elicitable with the mean. The revelation principle provides an explicit link between the class of strictly consistent scoring functions for the first two moments which are of Bregman-type (Fissler and Ziegel, 2016, Proposition 4.4) and the respective class for mean and variance.

As the pair (mean, variance) has of a non-elicitable component, it fails to be componentwise order-sensitive (Lemma B.2) and therefore, it is also not metrically order-sensitive. *A priori*, order-sensitivity on line segments is not ruled out. Corollary 3.16 implies that any strictly consistent scoring function for the pair of the first and second moment is order-sensitive on line segments. Even though the bijection connecting (mean, variance) with the pair of the first two moments is not linear, the following proposition gives necessary and sufficient conditions for scoring functions to be order-sensitive on line segments for (mean, variance). Example 3.19 shows the existence of order-sensitive scoring functions on line segments for (mean, variance).

Proposition 3.18. *Let \mathcal{F} be a class of distributions on \mathbb{R} with finite second moments such that the functional $T = (\text{mean}, \text{variance}): \mathcal{F} \rightarrow \mathbf{A}$ is surjective on $\mathbf{A} = \mathbb{R} \times (0, \infty)$. Let Assumptions (F1) and (V1) be satisfied with the strict \mathcal{F} -identification function $V: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}^2$, $V(x_1, x_2, y) = (x_1 - y, x_2 + x_1^2 - y^2)^\top$. Let $S: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$ be a scoring function that is (jointly) continuous and for any $y \in \mathbb{R}$, the function $\mathbf{A} \ni x \mapsto S(x, y)$ be twice continuously differentiable. Then S is \mathcal{F} -order-sensitive on line segments for T if and only if S is of the form*

$$S(x_1, x_2, y) = -\phi(x_1, x_2 + x_1^2) + \nabla \phi(x_1, x_2 + x_1^2) \begin{pmatrix} x_1 - y \\ x_2 + x_1^2 - y^2 \end{pmatrix} + a(y), \quad (3.8)$$

where $a: \mathbb{R} \rightarrow \mathbb{R}$ is some \mathcal{F} -integrable function and $\phi: \mathbf{A}' \rightarrow \mathbb{R}$, $\mathbf{A}' = \{(x_1, x_2 + x_1^2) \in \mathbb{R}^2 \mid x \in \mathbf{A}\} = \{(m_1, m_2) \in \mathbb{R}^2 \mid m_1^2 < m_2\}$, is a convex, three times continuously differentiable function such that the second order partial derivatives $\phi_{ij} := \partial_i \partial_j \phi$ satisfy

$$\phi_{12}(m_1, m_2) = -2m_1 \phi_{22}(m_1, m_2) \quad (3.9)$$

$$\phi_{11}(m_1, m_2) \geq (m_2 + 3m_1^2) \phi_{22}(m_1, m_2) \quad (3.10)$$

for all $(m_1, m_2) \in \mathbf{A}'$.

Example 3.19. An example for a class of strictly convex C^3 -function $\phi: \mathbf{A}' \rightarrow \mathbb{R}$ satisfying (3.9) and (3.10) with equality is given by

$$\phi(m_1, m_2) = (m_2 - m_1^2)^{-1} + b_1 m_1 + b_2 m_2 + b_3, \quad b_1, b_2, b_3 \in \mathbb{R}.$$

For the case $b_1 = b_2 = b_3 = 0$, the resulting scoring function of the form at (3.8) is

$$S(x_1, x_2, y) = x_2^{-2} (x_1^2 - 2x_2 - 2x_1 y + y^2). \quad (3.11)$$

Interestingly, this results not only in an order-sensitive scoring function on line segments for the pair (mean, variance), but it is also a mixed positively homogeneous scoring function of degree -2 ; see Section 4.2.

3.4.2. The pair (Value at Risk, Expected Shortfall)

Value at Risk (VaR) and Expected Shortfall (ES) are popular risk measures in banking and insurance. For a financial position Y with distribution F and a level $\alpha \in (0, 1)$, they are defined as

$$\begin{aligned}\text{VaR}_\alpha(F) &:= F^{-1}(\alpha) := \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}, \\ \text{ES}_\alpha(F) &:= \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\beta(F) \, d\beta \\ &= \frac{1}{\alpha} \mathbb{E}_F[Y \mathbb{1}\{Y \leq \text{VaR}_\alpha(F)\}] + \frac{1}{\alpha} \text{VaR}_\alpha(F)(\alpha - F(\text{VaR}_\alpha(F))).\end{aligned}$$

Note that if F is continuous at $\text{VaR}_\alpha(F)$, that means, if $F(\text{VaR}_\alpha(F)) = \alpha$, one can write $\text{ES}_\alpha(F)$ equivalently as $\mathbb{E}_F[Y | Y \leq \text{VaR}_\alpha(F)]$. Our sign convention implies that risky positions yield large *negative* values of VaR_α or ES_α . Intuitively, VaR_α gives the worst loss out of the best $(1 - \alpha) \times 100\%$ of all cases, whereas ES_α gives the average loss given one exceeds VaR_α . Merits and pitfalls of these two important risk measures are discussed in Embrechts et al. (2014); Embrechts and Hofert (2014) where numerous further references are given.

VaR_α , as a quantile, is elicitable under mild regularity conditions, whereas ES_α fails to be elicitable (Gneiting, 2011). However, recently it was shown in Fissler and Ziegel (2016, Theorem 5.2 and Corollary 5.5) that the pair $(\text{VaR}_\alpha, \text{ES}_\alpha)$ is elicitable and the class of strictly convex scoring functions was characterized to be of the form (3.12) (under the conditions of Osband's principle, Fissler and Ziegel (2016, Theorem 3.2, Corollary 3.3)). Note that the proof of Fissler and Ziegel (2016, Theorem 5.2(ii) and Corollary 5.5) is imprecise for the case that a distribution $F \in \mathcal{F}$ is not continuous at its α -quantile. Moreover, one needs to impose additional assumptions on the action domain \mathbf{A} which are satisfied, for example, if \mathbf{A} coincides with the maximal action domain $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq x_2\}$; see Fissler and Ziegel (2019) for details.

Proposition 3.20. *Let $\alpha \in (0, 1)$, \mathcal{F} be a class of continuously differentiable distribution functions on \mathbb{R} with finite first moments and unique α -quantiles. Let $\mathbf{A} \subseteq \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq x_2\}$ be convex. Define \mathbf{A}_2 as the projection of \mathbf{A} onto the second coordinate axis and let $S : \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$ be a scoring function of the form*

$$\begin{aligned}S(x_1, x_2, y) &= (\mathbb{1}\{y \leq x_1\} - \alpha)g(x_1) - \mathbb{1}\{y \leq x_1\}g(y) \\ &\quad + \phi'(x_2)\left(x_2 + (\mathbb{1}\{y \leq x_1\} - \alpha)\frac{x_1}{\alpha} - \mathbb{1}\{y \leq x_1\}\frac{y}{\alpha}\right) - \phi(x_2),\end{aligned}\tag{3.12}$$

with $g : \mathbb{R} \rightarrow \mathbb{R}$ differentiable and increasing and $\phi : \mathbf{A}_2 \rightarrow \mathbb{R}$ twice differentiable, and $\phi' > 0, \phi'' > 0$. If

$$\phi'(x) + (x - z)\phi''(x) \geq 0, \quad \text{for all } x, z \in \mathbf{A}_2,\tag{3.13}$$

then S is strictly \mathcal{F} -order-sensitive on line segments for $(\text{VaR}_\alpha, \text{ES}_\alpha)$.

One might wonder if Proposition 3.20 establishes an alternative set of conditions for strict consistency of scoring functions for $(\text{VaR}_\alpha, \text{ES}_\alpha)$ different from the ones introduced in Fissler and Ziegel (2019, Proposition 2). Indeed, this is the case since strict order-sensitivity on line segments implies strict consistency. However, it is not the condition at (3.13) which is essential for the strict consistency, but rather the condition that g be increasing and $\phi' > 0$, and $\phi'' > 0$.

Example 3.21. Consider the action domain $\mathbf{A} = \{x \in \mathbb{R}^2: x_1 \geq x_2, x_2 < 0\}$, so $\mathbf{A}_2 = (-\infty, 0)$. For all ϕ in the family $\{\phi_b: (-\infty, 0) \rightarrow \mathbb{R}: b \in (0, 1]\}$ where $\phi_1(x) = -\log(|x|)$, $x < 0$ and for $b \in (0, 1)$

$$\phi_b(x) = \frac{1}{b-1}|x|^{1-b}, \quad x < 0,$$

condition (3.13) is satisfied.

A strict \mathcal{F} -identification function $V: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ for a functional $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}$ is oriented for T if

$$\bar{V}(x, F) > 0 \iff x > T(F) \quad (3.14)$$

for all $F \in \mathcal{F}$, $x \in \mathbf{A}$ (Lambert, Pennock and Shoham, 2008; Steinwart et al., 2014). One possible generalization of orientation for higher-dimensional functionals is the following. Let $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ be a functional with a strict \mathcal{F} -identification function $V: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}^k$. Then V is called an *oriented* strict \mathcal{F} -identification function for T if

$$v^\top \bar{V}(T(F) + sv, F) > 0 \iff s > 0$$

for all $v \in \mathbb{S}^{k-1} := \{x \in \mathbb{R}^k: \|x\| = 1\}$, for all $F \in \mathcal{F}$ and for all $s \in \mathbb{R}$ such that $T(F) + sv \in \mathbf{A}$.

Our notion of orientation differs from the one proposed by Frongillo and Kash (2015a). In contrast to their definition, our definition is *per se* independent of a (possibly non-existing) strictly consistent scoring function for T . Moreover, whereas their definition has connections to the convexity of the expected score, our definition shows strong ties to order-sensitivity on line segments.

If the gradient of an expected score induces an oriented identification function, then the scoring function is strictly order-sensitive on line segments, and vice versa. However, the existence of an oriented identification function is not sufficient for the existence of a strictly order-sensitive scoring function on line segments. The reason is that – due to integrability conditions – the identification function is not necessarily the gradient of some (scoring) function.

4. Equivariant functionals and order-preserving scoring functions

Many statistical functionals have an invariance or equivariance property. For example, the mean is a linear functional, and hence, it is equivariant under linear transformations. So $\mathbb{E}[\varphi(X)] = \varphi(\mathbb{E}[X])$ for any random variable X and

any linear map $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ (of course, the same is true for the higher-dimensional setting). On the other hand, the variance is invariant under translations, that is $\text{Var}(X - c) = \text{Var}(X)$ for any $c \in \mathbb{R}$, but scales quadratically, so $\text{Var}(\lambda X) = \lambda^2 \text{Var}(X)$ for any $\lambda \in \mathbb{R}$. The next definition strives to formalize such notions.

Definition 4.1 (π -equivariance). Let \mathcal{F} be a class of probability distributions on \mathcal{O} and \mathcal{A} be an action domain. Let Φ be a group of bijective transformations $\varphi: \mathcal{O} \rightarrow \mathcal{O}$, Φ^* a group of bijective transformations $\varphi^*: \mathcal{A} \rightarrow \mathcal{A}$, and $\pi: \Phi \rightarrow \Phi^*$ be a map. A functional $T: \mathcal{F} \rightarrow \mathcal{A}$ is π -equivariant if for all $\varphi \in \Phi$

$$T(\mathcal{L}(\varphi(Y))) = (\pi\varphi)(T(\mathcal{L}(Y)))$$

for all random variables Y such that $\mathcal{L}(Y) \in \mathcal{F}$.

- Example 4.2.** (i) For $\mathcal{A} = \mathcal{O} = \mathbb{R}$, the mean functional is π -equivariant for $\Phi = \Phi^* = \{x \mapsto x + c, c \in \mathbb{R}\}$ the translation group and π the identity map, or for $\Phi = \Phi^* = \{x \mapsto \lambda x, \lambda \in \mathbb{R} \setminus \{0\}\}$ the multiplicative group and again π the identity map.
- (ii) For $\mathcal{A} = \mathcal{O} = \mathbb{R}^k$, the multivariate mean functional is π -equivariant for $\Phi = \Phi^* = \{x \mapsto x + c, c \in \mathbb{R}^k\}$ the translation group and π the identity map.
- (iii) For $\mathcal{A} = \mathcal{O} = \mathbb{R}$, Value at Risk at level α , Expected Shortfall at level α and the τ -expectile are π -equivariant for $\Phi = \Phi^* = \{x \mapsto x + c, c \in \mathbb{R}\}$ the translation group and π the identity map, or for $\Phi = \Phi^* = \{x \mapsto \lambda x, \lambda > 0\}$ the multiplicative group and again π the identity map.
- (iv) For $\mathcal{A} = [0, \infty)$ and $\mathcal{O} = \mathbb{R}$, the variance is π -equivariant for $\Phi = \{x \mapsto x + c, c \in \mathbb{R}\}$ the translation group and $\Phi^* = \{\text{id}_{\mathcal{A}}\}$ the trivial group consisting only of the identity on \mathcal{A} , such that π is the constant map.
- (v) For $\mathcal{A} = [0, \infty)$ and $\mathcal{O} = \mathbb{R}$, the variance is π -equivariant for $\Phi = \Phi^* = \{x \mapsto \lambda x, \lambda \in \mathbb{R} \setminus \{0\}\}$ the multiplicative group, and $\pi((x \mapsto \lambda x)) = (x \mapsto \lambda^2 x)$.
- (vi) Let $\mathcal{A} = \mathbb{R}^k$, $\mathcal{O} = \mathbb{R}$ and T be the functional whose m th component is the m th moment. Then T is π -equivariant with $\Phi = \{y \mapsto \lambda y, \lambda \in \mathbb{R} \setminus \{0\}\}$, $\Phi^* = \{x \mapsto (\lambda^m x_m)_{m=1}^k, \lambda \in \mathbb{R} \setminus \{0\}\}$, and $\pi((y \mapsto \lambda y)) = (x \mapsto (\lambda^m x_m)_{m=1}^k)$.

If a functional T is elicitable, π -equivariance can also be expressed in terms of strictly consistent scoring functions; see also Gneiting (2011, p. 750).

Lemma 4.3. Let $S: \mathcal{A} \times \mathcal{O} \rightarrow \mathbb{R}$ be a strictly \mathcal{F} -consistent scoring function for a functional $T: \mathcal{F} \rightarrow \mathcal{A}$ and let $\pi: \Phi \rightarrow \Phi^*$. Then, T is π -equivariant if and only if for all $\varphi \in \Phi$

$$\arg \min_{x \in \mathcal{A}} \bar{S}((\pi\varphi)(x), \mathcal{L}(\varphi(Y))) = \arg \min_{x \in \mathcal{A}} \bar{S}(x, \mathcal{L}(Y))$$

for all random variables Y such that $\mathcal{L}(Y) \in \mathcal{F}$.

The proof of Lemma 4.3 is direct. It implies that the scoring function

$$S_{\pi, \varphi}: \mathcal{A} \times \mathcal{O} \rightarrow \mathbb{R}, \quad (x, y) \mapsto S_{\pi, \varphi}(x, y) = S((\pi\varphi)(x), \varphi(y)) \quad (4.1)$$

is also strictly \mathcal{F} -consistent for T . Similarly to the motivation of order-sensitivity of scoring functions, for fixed $\pi: \Phi \rightarrow \Phi^*$, it is a natural requirement on a scoring function S that for all $\varphi \in \Phi$ the ranking of *any two* forecasts is the same in terms of S and in terms of $S_{\pi, \varphi}$.

Definition 4.4 (π -order-preserving). Let $\pi: \Phi \rightarrow \Phi^*$. A scoring function $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ is π -order-preserving with respect to \mathcal{F} if for all $\varphi \in \Phi$ one has

$$\operatorname{sgn}(\bar{S}(x, F) - \bar{S}(x', F)) = \operatorname{sgn}(\bar{S}_{\pi, \varphi}(x, F) - \bar{S}_{\pi, \varphi}(x', F))$$

for all $F \in \mathcal{F}$ and for all $x, x' \in \mathbf{A}$, where $S_{\pi, \varphi}$ is defined at (4.1). S is *linearly π -order-preserving* if for all $\varphi \in \Phi$ and for all $x, x' \in \mathbf{A}$ there is a $\lambda > 0$ such that

$$\lambda(S(x, y) - S(x', y)) = S_{\pi, \varphi}(x, y) - S_{\pi, \varphi}(x', y) \quad (4.2)$$

for all $y \in \mathbf{O}$. If S is linearly π -order-preserving with a $\lambda > 0$ independent of $x, x' \in \mathbf{A}$, then we call S *uniformly linearly π -order-preserving*.

The following lemma is immediate.

Lemma 4.5. *Let $\pi: \Phi \rightarrow \Phi^*$. If a scoring function $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ is linearly π -order-preserving, it is π -order-preserving with respect to any class \mathcal{F} of probability distributions on \mathbf{O} .*

The two practically most relevant examples of uniform linear π -order preservingness are translation invariance and positive homogeneity of scoring functions, or, to be more precise, of score differences. They are described in the two subsequent subsections.

4.1. Translation invariance

Consider a translation equivariant functional such as the mean treated in Example 4.2 (ii). Then, a scoring function $S: \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}$ is said to have translation invariant score differences if it is uniformly linearly π -equivariant with $\lambda = 1$ for all $\varphi \in \Phi$. In formulae, we require S to satisfy

$$S(x - z, y - z) - S(x' - z, y - z) = S(x, y) - S(x', y) \quad (4.3)$$

for all $x, x', y, z \in \mathbb{R}^k$. Note that what is particularly appealing is that the action domain and the observation domain coincide and, in particular, have the same dimension. However, there are also other functionals such as vectors of different quantiles or expectiles, or the vector $(\operatorname{VaR}_\alpha, \operatorname{ES}_\alpha)$ satisfying properties one can naturally call translation equivariant, but that have the drawback that $\mathbf{A} \neq \mathbf{O}$ (typically, \mathbf{O} is of lower dimension than \mathbf{A}). Then, translation invariance means that the score is invariant under a simultaneous translation of the observation and the forecast along respective linear subspaces of \mathbf{A} and \mathbf{O} .

Let $\mathbf{A} \subseteq \mathbb{R}^k$, $\mathbf{O} = \mathbb{R}^d$ and $m \in \{1, \dots, \min\{k, d\}\}$. Let $M_{\mathbf{O}} \in \mathbb{R}^{d \times m}$ and $M_{\mathbf{A}} \in \mathbb{R}^{k \times m}$ be two matrices with rank m . Define the transformation groups

$$\begin{aligned} \Phi &:= \Phi_{M_{\mathbf{O}}} := \{y \mapsto y - M_{\mathbf{O}}z, z \in \mathbb{R}^m\}, \\ \Phi^* &:= \Phi_{M_{\mathbf{A}}}^* := \{x \mapsto x - M_{\mathbf{A}}z, z \in \mathbb{R}^m\}, \end{aligned}$$

where we impose that $x - M_A z \in A$ for all $x \in A$, $z \in \mathbb{R}^m$. Then, the map $\pi = \pi_{M_O, M_A}: \Phi_{M_O} \rightarrow \Phi_{M_A}^*$ naturally induced by M_O and M_A is given as

$$\pi_{M_O, M_A}((y \mapsto y - M_O z)) = (x \mapsto x - M_A z).$$

We say that a functional $T: \mathcal{F} \rightarrow \mathbb{R}^k$ is linearly equivariant if there are such matrices M_O, M_A such that T is π_{M_O, M_A} -equivariant.

Example 4.6. (i) Let $O = \mathbb{R}$, $A = \{(x_1, x_2) \in \mathbb{R}^2: x_2 \leq x_1\}$ and $T = (\text{VaR}_\alpha, \text{ES}_\alpha)$ with some generic \mathcal{F} . Then T is π_{M_O, M_A} -equivariant with $M_O = \text{id}_{\mathbb{R}}$ and $M_A = (1, 1)^\top$.
(ii) Let $O = \mathbb{R}$, $A = \mathbb{R} \times [0, \infty)$ and $T = (\text{mean}, \text{variance}): \mathcal{F} \rightarrow A$ where all $F \in \mathcal{F}$ have finite second moments. Then T is π_{M_O, M_A} -equivariant with $M_O = \text{id}_{\mathbb{R}}$ and $M_A = (1, 0)^\top$.
(iii) Let $O = \mathbb{R}$, $A = \mathbb{R}^k$ and T be a vector of k different quantiles. Let $M \in \mathbb{R}^{k \times k}$ have rank at least 1 and consider the functional $T_M = M(T)$. Then T_M is π_{M_O, M_A} -equivariant with $M_O = \text{id}_{\mathbb{R}}$ and $M_A = M(1, \dots, 1)^\top$.

Adopting this notion, we say that a scoring function $S: A \times \mathbb{R}^d \rightarrow \mathbb{R}$ is linearly (M_O, M_A) -invariant for two matrices $M_O \in \mathbb{R}^{d \times m}$, $M_A \in \mathbb{R}^{k \times m}$ with $\text{rank}(M_O) = \text{rank}(M_A) = m \in \{1, \dots, \min\{k, d\}\}$ if

$$S(x - M_A z, y - M_O z) = S(x, y)$$

for all $x \in A$, $y \in \mathbb{R}^d$, $z \in \mathbb{R}^m$. Similarly, we will speak about linearly (M_O, M_A) -invariant identification functions and score differences.

Given a certain functional $T: \mathcal{F} \rightarrow \mathbb{R}^k$ and some $M_O \in \mathbb{R}^{d \times m}$, $M_A \in \mathbb{R}^{k \times m}$ with $\text{rank}(M_O) = \text{rank}(M_A) = m \in \{1, \dots, \min\{k, d\}\}$, one can wonder about the class of strictly consistent scoring functions that are linearly (M_O, M_A) -invariant. Clearly, with respect to Lemma 4.3 and Lemma 4.5, this class is empty if the functional T is not π_{M_O, M_A} -equivariant. In the situation that $A = O = \mathbb{R}^k$ and $M_O = M_A = \text{id}_{\mathbb{R}^k}$ the following proposition characterizes the gradients of linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant strictly consistent scoring function (if such scoring functions exist).

Proposition 4.7. *Let $T: \mathcal{F} \rightarrow \mathbb{R}^k$ be a surjective, identifiable functional with a linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant strict \mathcal{F} -identification function $V: \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^k$. Then, the following assertions hold.*

- (i) *T is $\pi_{\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k}}$ -equivariant.*
- (ii) *Assume there is a strictly \mathcal{F} -consistent scoring function $S: \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}$ for T with linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant score differences. Then, under Assumptions (V1) and (S1), there is a constant matrix $h \in \mathbb{R}^{k \times k}$ such that*

$$\nabla \bar{S}(x, F) = h \bar{V}(x, F) \quad (4.4)$$

for all $x \in \mathbb{R}^k$ and for all $F \in \mathcal{F}$.

Using Fissler and Ziegel (2016, Proposition 4.4) one can establish the converse of Proposition 4.7: If V is a linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant strict \mathcal{F} -identification

function, then (4.4) implies that S has linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant score differences. The following lemma shows how to normalize scores with translation invariant score differences to obtain a translation invariant score.

Lemma 4.8. *Let $S: \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}$ a strictly \mathcal{F} -consistent scoring function for $T: \mathcal{F} \rightarrow \mathbb{R}^k$ with linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant score differences. If for all $y \in \mathbb{R}^k$, the point measures δ_y are in \mathcal{F} and the function $y \mapsto S(T(\delta_y), y)$ is \mathcal{F} -integrable, then*

$$S_0(x, y) := S(x, y) - S(T(\delta_y), y) \quad (4.5)$$

is a linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant, non-negative, strictly \mathcal{F} -consistent scoring function for T .

In case of the mean functional on \mathbb{R} , Proposition 4.7 has already been shown by Savage (1971) who showed that the squared loss is the only strictly consistent scoring function for the mean that is of prediction error form, up to equivalence.¹³ Furthermore it implies that general τ -expectiles and α -quantiles have essentially one linearly $(\text{id}_{\mathbb{R}}, \text{id}_{\mathbb{R}})$ -invariant strictly consistent scoring function only, namely the canonical choices $S_\tau(x, y) = |\mathbb{1}\{y \leq x\} - \tau|(x - y)^2$ and $S_\alpha(x, y) = (\mathbb{1}\{y \leq x\} - \alpha)(x - y)$.

The uniqueness – up to equivalence – disappears for $k > 1$. For example, for the the 2-dimensional mean functional, the previous results yield that any scoring function $S: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ of the form

$$S(x, y) = \frac{h_{11}}{2}(x_1 - y_1)^2 + \frac{h_{22}}{2}(x_2 - y_2)^2 + h_{12}y_2(y_1 - x_1) + h_{12}x_2(x_1 - y_1)$$

is strictly consistent for the 2-dimensional mean functional and it is linearly $(\text{id}_{\mathbb{R}^2}, \text{id}_{\mathbb{R}^2})$ -invariant, for any $h_{11} > 0$ and $h_{11}h_{22} - h_{12}^2 > 0$.

Due to the additive separability of strictly consistent scoring functions for vectors consisting of different quantiles and expectiles (Fissler and Ziegel, 2016, Proposition 4.2), strictly consistent scoring functions that are linearly $(\text{id}_{\mathbb{R}}, \text{id}_{\mathbb{R}^k})$ -invariant for these vectors are not unique. However, the only flexibility in that class consists in choosing different weights for the respective summands of the scores.

The pair (mean, variance) is a π_{M_O, M_A} -equivariant functional with M_O and M_A as in Example 4.6(ii). Curiously, it has a linearly (M_O, M_A) -invariant identification function $V(x_1, x_2, y) = (x_1 - y, x_2 - (x_1 - y)^2)^\top$ but does not possess a strictly consistent linearly (M_O, M_A) -invariant scoring function.

Proposition 4.9. *Let \mathcal{F} be a class of distributions on \mathbb{R} with finite second moments such that the functional $T = (\text{mean}, \text{variance}): \mathcal{F} \rightarrow \mathbf{A}$ is surjective on $\mathbf{A} = \mathbb{R} \times I$, where $I \subseteq [0, \infty)$ is an interval. Let Assumptions (F1) and (V1) be satisfied with the strict \mathcal{F} -identification functions $V: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}^2$, $V(x_1, x_2, y) = (x_1 - y, x_2 - (x_1 - y)^2)^\top$ and $V^*: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}^2$, $V^*(x_1, x_2, y) = (x_1 - y, x_2 + x_1^2 - y^2)^\top$. Let $S: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$ be a \mathcal{F} -consistent scoring function for*

¹³That means that the scoring function is a function in $x - y$ only.

T that is (jointly) continuous, and for any $y \in \mathbb{R}$, the function $A \ni x \mapsto S(x, y)$ is twice continuously differentiable. If S has linearly (M_O, M_A) -invariant score differences, then there is a $\lambda \geq 0$ and an \mathcal{F} -integrable functional $a: \mathbb{R} \rightarrow \mathbb{R}$ such that

$$S(x_1, x_2, y) = \lambda(x_1 - y)^2 + a(y).$$

In particular, S cannot be strictly \mathcal{F} -consistent for T .

The functional $(\text{VaR}_\alpha, \text{ES}_\alpha)$, $\alpha \in (0, 1)$, is also a relevant π_{M_O, M_A} -equivariant functional with M_O and M_A as in Example 4.6(i). However, scoring functions with linearly (M_O, M_A) -invariant score differences only exist for restricted classes of distribution functions \mathcal{F} which may not be natural choices in risk management applications.

Proposition 4.10. *Let $\alpha \in (0, 1)$. Let \mathcal{F} be a class of distribution functions on \mathbb{R} with finite first moments and unique α -quantiles. Consider $T = (\text{VaR}_\alpha, \text{ES}_\alpha): \mathcal{F} \rightarrow \{(x_1, x_2) \in \mathbb{R}^2: x_2 \leq x_1\}$. Then, the following assertions hold:*

- (i) *Suppose there is some $c > 0$ such that*

$$\text{ES}_\alpha(F) + c > \text{VaR}_\alpha(F) \quad \text{for all } F \in \mathcal{F}. \quad (4.6)$$

That is, $T(\mathcal{F}) \subseteq A_c := \{(x_1, x_2) \in \mathbb{R}^2: x_2 \leq x_1 < x_2 + c\}$. Then, any scoring function $S: A_c \times \mathbb{R} \rightarrow \mathbb{R}$, which is equivalent to

$$S_c(x_1, x_2, y) = (\mathbb{1}\{y \leq x_1\} - \alpha)c(x_1 - y) + \alpha(x_2^2/2 + x_1^2/2 - x_1x_2) + \mathbb{1}\{y \leq x_1\}(-x_2(y - x_1) + y^2/2 - x_1^2/2), \quad (4.7)$$

is strictly \mathcal{F} -consistent for T and has linearly (M_O, M_A) -invariant score differences with $M_O = \text{id}_{\mathbb{R}}$, $M_A = (1, 1)^\top$.

- (ii) *Under the conditions of Fissler and Ziegel (2016, Theorem 5.2(iii)), there are strictly \mathcal{F} -consistent scoring functions for T with linearly (M_O, M_A) -invariant score differences if and only if there is some $c > 0$ such that (4.6) holds. Then, any such scoring function is necessarily equivalent to S_d defined at (4.7) almost everywhere, with $d \geq c$.*

The scoring function S_c has a close relationship to the class of scoring functions S^W proposed in Acerbi and Szekely (2014); see Fissler and Ziegel (2016, Equation (5.6)). Indeed, $S_c(x_1, x_2, y) = c(\mathbb{1}\{y \leq x_1\} - \alpha)(x - y) + S^W(x_1, x_2, y)$ with $W = 1$. That means it is the sum of the standard α -pinball loss for VaR_α – which is translation invariant – and S^1 . In the same flavor, the condition at (4.6) is similar to the one at Fissler and Ziegel (2016, Equation (5.7)). Since $\text{ES}_\alpha \leq \text{VaR}_\alpha$, the maximal action domain where S_c is strictly consistent is the stripe $A_c = \{(x_1, x_2) \in \mathbb{R}^2: x_2 \leq x_1 < x_2 + c\}$. Of course, by letting $c \rightarrow \infty$, one obtains the maximal sensible action domain $\{(x_1, x_2) \in \mathbb{R}^2: x_1 \geq x_2\}$ for the pair $(\text{VaR}_\alpha, \text{ES}_\alpha)$. However, considering the properly normalized version S_c/c , this converges to a strictly consistent scoring function for VaR_α as $c \rightarrow \infty$, but

which is independent of the forecast for ES_α . Hence, there is a caveat concerning the tradeoff between the size of the action domain and the sensitivity in the ES-forecast. This might cast doubt on the usage of scoring functions with translation invariant score differences for $(\text{VaR}_\alpha, \text{ES}_\alpha)$ in general.

Interestingly, the scoring function S_c at (4.7) has positively homogeneous score differences if and only if $c = 0$. However, $\mathbf{A}_0 = \emptyset$, which means that the requirement of translation invariance and homogeneity for score differences are mutually exclusive in case of strictly consistent scoring functions for $(\text{VaR}_\alpha, \text{ES}_\alpha)$.

4.2. Homogeneity

If one is interested in a positively homogeneous functional of degree one such as the mean, expectiles, quantiles, or ES, a scoring function $S: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is said to have positively homogeneous score differences of degree $b \in \mathbb{R}$ for this functional if the scoring function is uniformly linearly π -equivariant with $\Phi = \{\mathbb{R} \ni x \mapsto cx \in \mathbb{R}, c > 0\}$ the multiplicative group, π the identity on Φ and $\lambda = c^b$ in (4.2). This means that S needs to satisfy

$$S(cx, cy) - S(cz, cy) = c^b(S(x, y) - S(z, y)) \quad (4.8)$$

for all $x, z, y \in \mathbb{R}$ and $c > 0$. Since positive homogeneity of score differences is equivalent to invariance of forecast rankings under a change of unit, it has been argued that it is important in financial applications (Acerbi and Szekely, 2014). Nolde and Ziegel (2017) give a characterization of scoring functions with positively homogeneous score differences for many risk measures of applied interest, such as VaR / quantiles, expectiles, and the pair (VaR, ES); cf. Patton (2011) for results concerning the mean functional.

If the functional T is vector-valued, the degree of homogeneity can be different in the respective components, e.g. in case of the pair (mean, variance) or the vector consisting of the first k moments; cf. Example 4.2(vi). One can denote this property by *mixed positive homogeneity*, which means in case of the vector of the first k moments that

$$T(\mathcal{L}(cY)) = \Lambda(c)T(\mathcal{L}(Y)) \quad (4.9)$$

for all $c > 0$, where $\Lambda(c)$ is the $k \times k$ -diagonal matrix with diagonal elements c, c^2, \dots, c^k .¹⁴ In this situation, an interesting instant for uniformly linearly π -order-preserving scoring functions $S: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$ are those with *mixed positively homogeneous* score differences of degree $b \in \mathbb{R}$. That is,

$$S(\Lambda(c)x, cy) - S(\Lambda(c)z, cy) = c^b(S(xy) - S(z, y)) \quad (4.10)$$

for all $x, z \in \mathbf{A}$, $y \in \mathbb{R}$, and for all $c > 0$. With $k = 2$, corresponding assertions hold for the pair (mean, variance) and the respective scoring functions.

¹⁴Of course, we tacitly assume that for all $x \in \mathbf{A}$ and for all $c > 0$, we have $\Lambda(c)x \in \mathbf{A}$.

Proposition 4.11. *Let $\mathbf{A} \subseteq \mathbb{R}^k$ such that $\Lambda(c)x \in \mathbf{A}$ for all $c > 0$, $x \in \mathbf{A}$. Let $S: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$ be a consistent scoring function for the vector of the first k moments of the form*

$$S(x, y) = -\phi(x) + \nabla\phi(x) (x - (y, y^2, \dots, y^k)^\top) + a(y), \quad (4.11)$$

where $\phi: \mathbf{A} \rightarrow \mathbb{R}$ is convex and differentiable with gradient $\nabla\phi$ (considered as a row vector). Then S has mixed positively homogeneous score differences of degree $b \in \mathbb{R}$ if and only if for all $c > 0$ the map

$$x \mapsto \nabla\phi(\Lambda(c)x)\Lambda(c) - c^b \nabla\phi(x) \quad (4.12)$$

is constant.

Recall that the scoring functions of the form at (4.11) are essentially all consistent scoring functions for the vector of different moments (Fissler and Ziegel, 2016, Proposition 4.4). Using Proposition 4.11 it is straight forward to derive consistent scoring functions for (mean, variance) with mixed positively homogeneous score differences.

Corollary 4.12. *Let \mathcal{F} be a class of distributions on \mathbb{R} with finite second moments such that the functional $T = (\text{mean}, \text{variance}): \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R} \times [0, \infty)$ is surjective, where for all $(x_1, x_2) \in \mathbf{A}$ and $c > 0$, $(cx_1, c^2x_2) \in \mathbf{A}$. Let Assumptions (F1) and (V1) be satisfied with the strict \mathcal{F} -identification function $V: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}^2$, $V(x_1, x_2, y) = (x_1 - y, x_2 + x_1^2 - y^2)^\top$. Let $S: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$ be a strictly \mathcal{F} -consistent scoring function for T that is (jointly) continuous and for any $y \in \mathbb{R}$, the function $\mathbf{A} \ni x \mapsto S(x, y)$ be twice continuously differentiable. Then S has mixed positively homogeneous score differences of degree $b \in \mathbb{R}$ if and only if*

$$S(x_1, x_2, y) = -\phi(x_1, x_2 + x_1^2) + \nabla\phi(x_1, x_2 + x_1^2) \begin{pmatrix} x_1 - y \\ x_2 + x_1^2 - y^2 \end{pmatrix} + a(y), \quad (4.13)$$

where $\phi: \mathbf{A} \rightarrow \mathbb{R}$ is strictly convex, twice continuously differentiable, and moreover for all $c > 0$ the map

$$\mathbf{A} \ni (x_1, x_2) \mapsto \nabla\phi(cx_1, c^2x_2 + c^2x_1^2) \begin{pmatrix} c & 0 \\ 0 & c^2 \end{pmatrix} - c^b \nabla\phi(x_1, x_2 + x_1^2) \quad (4.14)$$

is constant.

It appears that the class of (strictly) convex functions ϕ satisfying (4.12) is rather flexible. One subclass is the class of additively separable functions ϕ . That is,

$$\phi(x) = \sum_{m=1}^k \phi_m(x_m), \quad (4.15)$$

where each ϕ_m needs to be convex and $x_m \mapsto c^m \phi'_m(c^m x_m) - c^b \phi'_m(x_m)$ constant. Reviewing Nolde and Ziegel (2017, Theorem 5) and restricting attention to the

case $A \subseteq (0, \infty)^k$, ϕ_m can be an element of the class $\Psi_{b/m}$, where Ψ_b consists of functions $\psi_b: (0, \infty) \rightarrow \mathbb{R}$ of the form

$$\psi_b(y) = \begin{cases} d_0 + d_1 y^b / (b(b-1)), & \text{for } b \in \mathbb{R} \setminus \{0, 1\} \\ d_0 + d_1 y \log(y) + d_2 y, & \text{for } b = 1 \\ d_0 - d_1 \log(y) + d_2 y, & \text{for } b = 0 \end{cases}$$

with constants $d_1 > 0$, $d_0, d_2 \in \mathbb{R}$. On the other hand, there are choices of ϕ not satisfying such an additive decomposition as in (4.15). One such example can be found in Example 3.19 for $b = -2$, and is of the form $\phi(x_1, x_2) = (x_2 - x_1^2)^{-1}$ for $x_2 > x_1^2$.

Appendix A: Assumptions

We present a list of assumptions used in this paper. For more details about their interpretations and implications, please see Fissler and Ziegel (2016) where they were originally introduced.

Assumption (V1). Let \mathcal{F} be a convex class of distribution functions on \mathbb{R} and assume that for every $x \in \text{int}(A)$ there are $F_1, \dots, F_{k+1} \in \mathcal{F}$ such that

$$0 \in \text{int} \left(\text{conv} \left(\{ \bar{V}(x, F_1), \dots, \bar{V}(x, F_{k+1}) \} \right) \right).$$

Note that if $V: A \times \mathbb{R} \rightarrow \mathbb{R}^k$ is a strict \mathcal{F} -identification function for $T: \mathcal{F} \rightarrow A$ which satisfies Assumption (V1), then for each $x \in \text{int}(A)$ there is an $F \in \mathcal{F}$ such that $T(F) = x$.

Assumption (V2). For every $F \in \mathcal{F}$, the function $\bar{V}(\cdot, F)$ is continuous.

Assumption (V3). For every $F \in \mathcal{F}$, the function $\bar{V}(\cdot, F)$ is continuously differentiable.

Assumption (V4). Let assumption (V3) hold. For all $r \in \{1, \dots, k\}$ and for all $t \in \text{int}(A) \cap T(\mathcal{F})$ there are $F_1, F_2 \in T^{-1}(\{t\})$ such that

$$\partial_l \bar{V}_l(t, F_1) = \partial_l \bar{V}_l(t, F_2) \quad \forall l \in \{1, \dots, k\} \setminus \{r\}, \quad \partial_r \bar{V}_r(t, F_1) \neq \partial_r \bar{V}_r(t, F_2).$$

Assumption (F1). For every $y \in \mathbb{R}$ there exists a sequence $(F_n)_{n \in \mathbb{N}}$ of distributions $F_n \in \mathcal{F}$ that converges weakly to the Dirac-measure δ_y such that the support of F_n is contained in a compact set K for all n .

Assumption (VS1). Suppose that the complement of the set

$$C := \{(x, y) \in A \times \mathbb{R} \mid V(x, \cdot) \text{ and } S(x, \cdot) \text{ are continuous at the point } y\}$$

has $(k + d)$ -dimensional Lebesgue measure zero.

Assumption (S1). For every $F \in \mathcal{F}$, the function $\bar{S}(\cdot, F): \mathbf{A} \rightarrow \mathbb{R}$, $x \mapsto \bar{S}(x, F)$, is continuously differentiable.

Assumption (S2). For every $F \in \mathcal{F}$, the function $\bar{S}(\cdot, F)$ is continuously differentiable and the gradient is locally Lipschitz continuous. Furthermore, $\bar{S}(\cdot, F)$ is twice continuously differentiable at $t = T(F) \in \text{int}(\mathbf{A})$.

Appendix B: Auxiliary results

Lemma B.1. Let \mathcal{F} be convex and $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ be identifiable with a strict \mathcal{F} -identification function $V: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}^k$. Then for any $F, G \in \mathcal{F}$, the path $\gamma: [0, 1] \rightarrow \mathbf{A}$, $\gamma(\lambda) = T(\lambda F + (1 - \lambda)G)$, is either constant or injective.

Proof. Let $F, G \in \mathcal{F}$ such that $t = T(F) = T(G)$. For any $\lambda \in [0, 1]$, one has $\bar{V}(t, \lambda F + (1 - \lambda)G) = \lambda \bar{V}(t, F) + (1 - \lambda) \bar{V}(t, G) = 0$. Since V is a strict \mathcal{F} -identification function for T , $t = \gamma(\lambda)$ for all $\lambda \in [0, 1]$.

Now let $T(F) \neq T(G)$ and let $0 \leq \lambda < \lambda' \leq 1$. Since V is a strict \mathcal{F} -identification function, $\bar{V}(T(F), G) \neq 0$ (and symmetrically $\bar{V}(T(G), F) \neq 0$). Assume that $\gamma(\lambda) = \gamma(\lambda')$. Define $H_\lambda = \lambda F + (1 - \lambda)G$, $H_{\lambda'} = \lambda' F + (1 - \lambda')G$. There are $\mu, \mu' \in \mathbb{R}$ such that $F = \mu H_\lambda + (1 - \mu)H_{\lambda'}$ and $G = \mu' H_\lambda + (1 - \mu')H_{\lambda'}$. Hence,

$$\bar{V}(\gamma(\lambda), F) = \mu \bar{V}(\gamma(\lambda), H_\lambda) + (1 - \mu) \bar{V}(\gamma(\lambda), H_{\lambda'}) = 0,$$

and similarly $\bar{V}(\gamma(\lambda), G) = 0$. Consequently, $T(F) = \gamma(\lambda) = T(G)$, which is a contradiction to the assumption that $T(F) \neq T(G)$. This implies that $\gamma(\lambda) \neq \gamma(\lambda')$. \square

Lemma B.2. Let $T = (T_1, \dots, T_k): \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ be a k -dimensional functional with components $T_m: \mathcal{F} \rightarrow \mathbf{A}_m \subseteq \mathbb{R}$ where $\mathbf{A} = \mathbf{A}_1 \times \dots \times \mathbf{A}_k$. If there is a strictly componentwise \mathcal{F} -order-sensitive scoring function $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ for T , then the components T_m , $m \in \{1, \dots, k\}$, are elicitable.

Proof. Fix $m \in \{1, \dots, k\}$. Let $F \in \mathcal{F}$ and $x, z \in \mathbf{A}$ such that $T_m(F) = x_m$, $x_i = z_i$ for all $i \neq m$ and $x_m \neq z_m$. Due to the strict componentwise \mathcal{F} -order-sensitivity of S this implies that $\bar{S}(x, F) < \bar{S}(z, F)$. This in turn means that for any $z = (z_1, \dots, z_k) \in \mathbf{A}$ the map $S_{m,z}: \mathbf{A}_m \times \mathbf{O} \rightarrow \mathbb{R}$,

$$(x_m, y) \mapsto S_{m,z}(x_m, y) := S(z_1, \dots, z_{m-1}, x_m, z_{m+1}, \dots, z_k, y) \quad (\text{B.1})$$

is a strictly \mathcal{F} -consistent scoring function for T_m . \square

Lemma B.3. Let \mathcal{F} be convex and $T: \mathcal{F} \rightarrow \mathbf{A} \subseteq \mathbb{R}^k$ be mixture-continuous and surjective. Let $S: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ be (strictly) \mathcal{F} -consistent for T . Then S is (strictly) metrically \mathcal{F} -order-sensitive for T relative to $\|\cdot\|$ if and only if for all $F \in \mathcal{F}$, $t = T(F)$ and $x, z \in \mathbf{A}$ we have the implication

$$\|x - t\| = \|z - t\| \implies \bar{S}(x, F) = \bar{S}(z, F). \quad (\text{B.2})$$

Proof. Let S be metrically \mathcal{F} -order sensitive for T relative to d . Let $F \in \mathcal{F}$, $t = T(F)$, $x, z \in \mathbf{A}$ such that $\|x - t\| = \|z - t\|$. Then we have both $\bar{S}(x, F) \leq \bar{S}(z, F)$ and $\bar{S}(z, F) \leq \bar{S}(x, F)$.

Assume that (B.2) holds and S is (strictly) \mathcal{F} -consistent. Let $F \in \mathcal{F}$ with $t = T(F)$ and $x, z \in \mathbf{A}$. Suppose that $\|x - t\| \leq \|z - t\|$. If $\|x - t\| = \|z - t\|$, (B.2) implies that $\bar{S}(x, F) = \bar{S}(z, F)$ and there is nothing to show. If $\|x - t\| < \|z - t\|$, we can apply Proposition 2.4. There is a continuous path $\gamma: [0, 1] \rightarrow \mathbf{A}$ such that $\gamma(0) = z$ and $\gamma(1) = t$, and the function $[0, 1] \ni \lambda \mapsto \bar{S}(\gamma(\lambda), F)$ is decreasing. Due to continuity there is a $\lambda' \in [0, 1]$ such that $\|\gamma(\lambda') - t\| = \|x - t\|$. Invoking (B.2) it holds that $\bar{S}(x, F) = \bar{S}(\gamma(\lambda'), F) \leq \bar{S}(z, F)$. If S is strictly \mathcal{F} -consistent then the latter inequality is strict. \square

Appendix C: Proofs

C.1. Proofs for Section 2

Proof of Proposition 2.2. Let $\mathcal{F}_0 \subseteq \mathcal{F}$ be convex such that $T(\mathcal{F}_0) \subseteq [-C, C]^k$ for some $C > 0$. Let $F, G \in \mathcal{F}_0$. Define $h_{F,G}: [-C, C]^k \times [0, 1] \rightarrow \mathbb{R}$ via

$$h_{F,G}(x, \lambda) = \bar{S}(x, (1 - \lambda)F + \lambda G) = (1 - \lambda)\bar{S}(x, F) + \lambda\bar{S}(x, G).$$

Then $h_{F,G}$ is jointly continuous, and due to the strict consistency

$$T((1 - \lambda)F + \lambda G) = \arg \min_{x \in [-C, C]^k} h_{F,G}(x, \lambda).$$

By virtue of the Berge Maximum Theorem (Aliprantis and Border, 2006, Theorem 17.31 and Lemma 17.6), the function $\lambda \mapsto \arg \min_{x \in [-C, C]^k} h_{F,G}(x, \lambda)$ is continuous. \square

Proof of Proposition 2.4. Let $F \in \mathcal{F}$, $t = T(F)$ and $x \neq t$. Then there is some $G \in \mathcal{F}$ with $x = T(G)$. Define $\gamma: [0, 1] \rightarrow \mathbf{A}$, $\lambda \mapsto T(\lambda F + (1 - \lambda)G)$. Clearly, $\gamma(0) = x$ and $\gamma(1) = t$. Due to the mixture-continuity of T , the path γ is also continuous. The rest follows along the lines of the proof of Nau (1985, Proposition 3). Let $0 \leq \lambda < \lambda' \leq 1$. If $\gamma(\lambda) = \gamma(\lambda')$, there is nothing to show. So assume that $\gamma(\lambda) \neq \gamma(\lambda')$. Define $H_\lambda = \lambda F + (1 - \lambda)G$, and $H_{\lambda'}$ analogously. Then, for $\mu := (\lambda' - \lambda)/(1 - \lambda) \in (0, 1]$, it holds that $H_{\lambda'} = \mu F + (1 - \mu)H_\lambda$. The strict consistency of S implies that

$$\begin{aligned} \mu \bar{S}(\gamma(\lambda'), F) + (1 - \mu) \bar{S}(\gamma(\lambda'), H_\lambda) &= \bar{S}(\gamma(\lambda'), H_{\lambda'}) \\ &< \bar{S}(\gamma(\lambda), H_{\lambda'}) = \mu \bar{S}(\gamma(\lambda), F) + (1 - \mu) \bar{S}(\gamma(\lambda), H_\lambda), \end{aligned}$$

which is equivalent to

$$\frac{1 - \mu}{\mu} (\bar{S}(\gamma(\lambda'), H_\lambda) - \bar{S}(\gamma(\lambda), H_\lambda)) < \bar{S}(\gamma(\lambda), F) - \bar{S}(\gamma(\lambda'), F).$$

By strict consistency of S , the left-hand side is non-negative yielding the assertion. \square

Proof of Proposition 2.6. Let $F \in \mathcal{F}$ with $t = T(F)$. Due to the strict \mathcal{F} -consistency of S , the expected score $\bar{S}(\cdot, F)$ has a local minimum at t . Assume there is another local minimum at some $x \neq t$. Then there is a distribution $G \in \mathcal{F}$ with $x = T(G)$. Consider the path $\gamma: [0, 1] \rightarrow \mathbf{A}$, $\lambda \mapsto T(\lambda F + (1-\lambda)G)$. Due to Proposition 2.4 the function $\lambda \mapsto \bar{S}(\gamma(\lambda), F)$ is decreasing and strictly decreasing when we move on the image of the path from x to t . Hence $\bar{S}(\cdot, F)$ cannot have a local minimum at $x = \gamma(0)$. \square

Proof of Proposition 2.8. Let $F \in \mathcal{F}$, $t = T(F)$ and $\varepsilon > 0$. Define

$$\delta := \min\{\bar{S}(z, F) - \bar{S}(t, F) : z \in \mathbf{A}, \|z - t\| = \varepsilon\}.$$

Due to the continuity of $\bar{S}(\cdot, F)$, the minimum is well-defined and, as a consequence of the strict \mathcal{F} -consistency of S for T , δ is positive. Let $x \in \mathbf{A}$. If $\|x - t\| = \varepsilon$, we have, by the definition of δ , that $\bar{S}(x, F) - \bar{S}(t, F) \geq \delta$. Assume that $\|x - t\| > \varepsilon$. Then there is a distribution $G \in \mathcal{F}$ with $T(G) = x$. Due to Proposition 2.4 there is a continuous path $\gamma: [0, 1] \rightarrow \mathbf{A}$ such that $\gamma(0) = x$, $\gamma(1) = t$ and such that $\bar{S}(\gamma(\lambda), F)$ is decreasing in λ . Moreover, if $\lambda < \lambda'$ such that $\gamma(\lambda) \neq \gamma(\lambda')$ it holds that $\bar{S}(\gamma(\lambda), F) > \bar{S}(\gamma(\lambda'), F)$. Due to the continuity of γ there is some $x' \in \gamma([0, 1])$ with $\|x' - t\| = \varepsilon$. Then we obtain $\bar{S}(x, F) - \bar{S}(t, F) > \bar{S}(x', F) - \bar{S}(t, F) \geq \delta$. \square

C.2. Proofs for Section 3

Proof of Proposition 3.6. Due to the fact that for fixed $y \in \mathbf{O}$, $V(x, y)$ is a polynomial in x , Assumption (V3) is automatically satisfied. Let $h: \text{int}(\mathbf{A}) \rightarrow \mathbb{R}^{k \times k}$ be the matrix-valued function given in Osband's principle; see Fissler and Ziegel (2016, Theorem 3.2). By Fissler and Ziegel (2016, Proposition 4.4(i)) we have that

$$\partial_l h_{rm}(x) = \partial_r h_{lm}(x), \quad h_{rl}(x) = h_{lr}(x) \quad (\text{C.1})$$

for all $r, l, m \in \{1, \dots, k\}$, $l \neq r$, where the first identity holds for almost all $x \in \text{int}(\mathbf{A})$ and the second identity for all $x \in \text{int}(\mathbf{A})$. Moreover, the matrix $(h_{rl}(x))_{l,r=1,\dots,k}$ is positive definite for all $x \in \text{int}(\mathbf{A})$. If we can show that $h_{lr} = 0$ for $l \neq r$, we can use the first part of (C.1) and deduce that for all $m \in \{1, \dots, k\}$ there are positive functions $g_m: \mathbf{A}'_m \rightarrow \mathbb{R}$, where $\mathbf{A}'_m = \{x_m \in \mathbb{R} : \exists(z_1, \dots, z_k) \in \text{int}(\mathbf{A}) \text{ and } z_m = x_m\}$, such that

$$h_{mm}(x_1, \dots, x_k) = g_m(x_m)$$

for all $(x_1, \dots, x_k) \in \text{int}(\mathbf{A})$. Then, we can conclude like in the proof of Fissler and Ziegel (2016, Proposition 4.2(ii)).¹⁵

¹⁵The arguments in Fissler and Ziegel (2016, Proposition 4.2(ii)) use Fissler and Ziegel (2016, Proposition 3.4). There is a flaw in the latter result which has been pointed out in Brehmer (2017). We present a corrected version of the result in the erratum Fissler and Ziegel (2019).

Fix $l, r \in \{1, \dots, k\}$ with $l \neq r$ and $F \in \mathcal{F}$ such that $T(F) \in \text{int}(\mathbf{A})$. Due to the strict \mathcal{F} -consistency of $S_{l,z}$ defined at (B.1) we have that

$$0 = \frac{d}{dx_l} \bar{S}_{l,z}(x_l, F) = \partial \bar{S}_{l,z}(x_l, F) = \partial_l \bar{S}(z_1, \dots, z_{l-1}, x_l, z_{l+1}, \dots, z_k, F)$$

whenever $x_l = T_l(F)$ and for all $z \in \text{int}(\mathbf{A})$. This means the map $\text{int}(\mathbf{A}) \ni z \mapsto \partial \bar{S}_{l,z}(T_l(F), F)$ is constantly 0. Hence, for all $x \in \text{int}(\mathbf{A})$

$$\partial_r \partial_l \bar{S}(x, F) = 0$$

whenever $x_l = T_l(F)$. Using the special form of V and Fissler and Ziegel (2016, Corollary 3.3), we have for $x = t = T(F)$ that

$$0 = \partial_r \partial_l \bar{S}(t, F) = h_{lr}(t) \partial_r \bar{V}_r(t, F) = h_{lr}(t) \bar{q}(F)$$

and by assumption $\bar{q}(F) > 0$. Using the surjectivity of T we obtain that $h_{lr}(t) = 0$ for all $t \in \text{int}(\mathbf{A})$, which ends the proof. \square

Proof of Proposition 3.7. We apply Osband's principle, that is, Fissler and Ziegel (2016, Theorem 3.2) to S . Consequently, there is a function $h: \text{int}(\mathbf{A}) \rightarrow \mathbb{R}$ such that

$$\frac{d}{dx} \bar{S}(x, F) = h(x) \bar{V}(x, F) \quad (\text{C.2})$$

for all $F \in \mathcal{F}$, $x \in \text{int}(\mathbf{A})$. Due to the strict \mathcal{F} -consistency of S and the orientation of V , it holds that $h \geq 0$. We show that actually $h > 0$. Applying Lemma B.3, one has that

$$\bar{S}(T(F) + x, F) = \bar{S}(T(F) - x, F) \quad (\text{C.3})$$

for all $F \in \mathcal{F}$, $x \in \mathbb{R}$ such that $T(F) + x, T(F) - x \in \text{int}(\mathbf{A})$. Hence, also the derivative with respect to x of the left-hand side of (C.3) must coincide with the derivative on the right-hand side. This yields, using (C.2),

$$h(T(F) + x) \bar{V}(T(F) + x, F) = -h(T(F) - x) \bar{V}(T(F) - x, F) \quad (\text{C.4})$$

for all $F \in \mathcal{F}$, $x \in \mathbb{R}$ such that $T(F) + x, T(F) - x \in \text{int}(\mathbf{A})$. Assume $h(z) = 0$ for some $z \in \text{int}(\mathbf{A})$. Then, by surjectivity of T and convexity of $\text{int}(\mathbf{A})$, for all $z' \in \text{int}(\mathbf{A}) \setminus \{z\}$ there exists an $F \in \mathcal{F}$ and $x \in \mathbb{R} \setminus \{0\}$ such that $z = T(F) + x$ and $z' = T(F) - x$. Since V is a *strict* \mathcal{F} -identification function for T , both $\bar{V}(T(F) + x, F) \neq 0$ and $\bar{V}(T(F) - x, F) \neq 0$. Hence, (C.4) implies that $h(z') = 0$. This implies that h identically vanishes on $\text{int}(\mathbf{A})$ which contradicts the strict \mathcal{F} -consistency of S .

Therefore, $V^*(x, y) := h(x)V(x, y)$ is an oriented strict \mathcal{F} -identification function for T . Applying Osband's principle to S^* , one obtains a function $h^*: \text{int}(\mathbf{A}) \rightarrow \mathbb{R}$ such that $d/(dx) \bar{S}^*(x, F) = h^*(x) \bar{V}^*(x, F)$ for all $F \in \mathcal{F}$, $x \in \mathbb{R}$ such that $T(F) + x, T(F) - x \in \text{int}(\mathbf{A})$. Due to the analogue of (C.3) for S^* and (C.4), one obtains

$$\begin{aligned} h^*(T(F) + x) \bar{V}^*(T(F) + x, F) &= -h^*(T(F) - x) \bar{V}^*(T(F) - x, F) \\ &= h^*(T(F) - x) \bar{V}^*(T(F) + x, F). \end{aligned}$$

for all $F \in \mathcal{F}$, $x \in \mathbb{R}$ with $T(F) + x, T(F) - x \in \text{int}(\mathbf{A})$. By a similar reasoning as above, one can deduce that h^* must be constant and positive. Now, the claim follows by Fissler and Ziegel (2016, Proposition 3.4); see Fissler and Ziegel (2019) for a correction. \square

Proof of Proposition 3.8. (i) Let $m \in \{1, \dots, k\}$, $F \in \mathcal{F}$ with $t = T(F) \in \text{int}(\mathbf{A})$. Let $\mu \in \mathbb{R}$ and $x, z \in \text{int}(\mathbf{A})$ with $x_i = z_i = t_i$ for all $i \neq m$ and with $x_m = t_m + \mu$ and $z_m = t_m - \mu$, such that $|x - t| = |z - t|$. Due to Lemma B.3 and due to the particular additive form of S , we have

$$\begin{aligned} 0 = \bar{S}(x, F) - \bar{S}(z, F) &= \bar{S}_m(x_m, F) - \bar{S}_m(z_m, F) \\ &= \bar{S}_m(t_m + \mu, F) - \bar{S}_m(t_m - \mu, F). \end{aligned}$$

Again with Lemma B.3 one obtains the assertion.

(ii) The only interesting direction is to assume that S^* is strictly metrically \mathcal{F} -order-sensitive (with respect to the same ℓ^p -norm as S). We will show that $\lambda_1 = \lambda_m$ for all $m \in \{2, \dots, k\}$. Let $F \in \mathcal{F}$, $t = T(F) \in \text{int}(\mathbf{A})$, $x, z \in \text{int}(\mathbf{A})$ with $\|x - t\|_p = \|z - t\|_p > 0$ and $x_i = z_i = t_i$ for all $i \in \{2, \dots, k\} \setminus \{m\}$. Moreover, let $x_1 \neq z_1 = t_1$. Due to Lemma B.3 we have that $\bar{S}(x, F) - \bar{S}(z, F) = \bar{S}^*(x, F) - \bar{S}^*(z, F) = 0$. Moreover,

$$\begin{aligned} 0 = \bar{S}(x, F) - \bar{S}(z, F) &= \sum_{i=1}^k \bar{S}_i(x_i, F) - \bar{S}_i(z_i, F) \\ &= \bar{S}_1(x_1, F) - \bar{S}_1(z_1, F) + \bar{S}_m(x_m, F) - \bar{S}_m(z_m, F). \end{aligned}$$

Setting $\varepsilon := \bar{S}_1(x_1, F) - \bar{S}_1(z_1, F) > 0$, one obtains with the same calculation

$$\begin{aligned} 0 &= \bar{S}^*(x, F) - \bar{S}^*(z, F) \\ &= \lambda_1 (\bar{S}_1(x_1, F) - \bar{S}_1(z_1, F)) + \lambda_m (\bar{S}_m(x_m, F) - \bar{S}_m(z_m, F)) \\ &= \varepsilon (\lambda_1 - \lambda_m). \end{aligned} \quad \square$$

Proof of Proposition 3.9. (i) We can apply Lemma B.3. Let $F \in \mathcal{F}$. Then

$$\mathbb{R} \ni x \mapsto \bar{S}(T(F) + x, F) = \frac{1}{2} \bar{q}(F) x^2 - \frac{1}{2} \frac{\bar{p}(F)^2}{\bar{q}(F)}$$

is an even function in x . Moreover, equivalence of scoring functions preserves (strict) metrical order-sensitivity.

(ii) The convexity of \mathbf{A} is implied by the mixture-continuity of T and the convexity of \mathcal{F} . Then, the claim follows with Proposition 3.7. \square

Proof of Proposition 3.10. To show (i) we apply again Lemma B.3. For any $F \in \mathcal{F}$, $x \in \mathbb{R}^k$, we have $\bar{S}(T(F) + x, F) = (1/2) \bar{q}(F) \|x\|_2^2 - 1/(2 \bar{q}(F)) \sum_{m=1}^k \bar{p}_m(F)^2$ which only depends on the ℓ^2 -norm of x .

We prove (ii) and (iii) together. Assume there is a scoring function S^* satisfying the conditions above, so in particular, it is strictly metrically \mathcal{F} -order-sensitive with respect to the ℓ^p -norm for $p \in [1, \infty)$. Invoking Lemma 3.5(i), S^*

is strictly componentwise \mathcal{F} -order-sensitive for T . Thanks to Proposition 3.6, S^* is additively separable. By Proposition 3.9(i), it is of the form

$$S^*(x_1, \dots, x_k) = \sum_{m=1}^k \lambda_m \left(\frac{1}{2} q(y) x_m^2 - p_m(y) x_m \right) + \sum_{m=1}^k a_m(y).$$

If $p = 2$, part (i) and Proposition 3.8(ii) yield that $\lambda_1 = \dots = \lambda_k$, and hence, S and S^* are equivalent. For $p \neq 2$, we obtain $\bar{S}(T(F) + x, F) = (1/2)\bar{q}(F) \sum_{m=1}^k \lambda_m x_m^2 - 1/(2\bar{q}(F)) \sum_{m=1}^k \bar{p}_m(F)^2$. It is not hard to see that there are always x, x' with $\|x\|_p = \|x'\|_p$ but $\bar{S}(T(F) + x, F) \neq \bar{S}(T(F) + x', F)$. \square

Proof of Proposition 3.11. Assume that there exists a strictly metrically \mathcal{F} -order-sensitive scoring function $S_\alpha: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ satisfying Assumption (S1). Due to Lemma B.3, for any $F \in \mathcal{F}$ and any $x \in \mathbb{R}$

$$\bar{S}_\alpha(T_\alpha(F) + x, F) = \bar{S}_\alpha(T_\alpha(F) - x, F).$$

Using Osband's principle (Fissler and Ziegel, 2016, Theorem 3.2) and taking the derivative with respect to x on both sides, this yields

$$h(T_\alpha(F) + x) \bar{V}_\alpha(T_\alpha(F) + x, F) = -h(T_\alpha(F) - x) \bar{V}_\alpha(T_\alpha(F) - x, F) \quad (\text{C.5})$$

for some positive function $h: \mathbb{R} \rightarrow \mathbb{R}$ (the fact that $h \geq 0$ follows from the strict consistency of S_α and the surjectivity of T_α , and $h > 0$ follows like in the proof of Proposition 3.7). Assume that $T_\alpha(F_0) = 0$. For $\lambda \in \mathbb{R}$, we have $T_\alpha(F_0(\cdot - \lambda)) = \lambda$. Therefore, (C.5) implies

$$\frac{h(\lambda + x)}{h(\lambda - x)} = -\frac{\bar{V}_\alpha(\lambda - x, F_0(\cdot - \lambda))}{\bar{V}_\alpha(\lambda + x, F_0(\cdot - \lambda))} = -\frac{F_0(-x) - \alpha}{F_0(x) - \alpha}. \quad (\text{C.6})$$

Setting $\lambda = \pm x$, one can see that $h(\pm\infty) := \lim_{x \rightarrow \pm\infty} h(x)$ exists and that $h(+\infty) = h(0)\alpha/(1 - \alpha)$, $h(-\infty) = h(0)(1 - \alpha)/\alpha$, hence, $h(+\infty)/h(-\infty) = 1$. On the other hand, for fixed $\lambda \in \mathbb{R}$, we obtain

$$\frac{h(+\infty)}{h(-\infty)} = \lim_{x \rightarrow \infty} \frac{h(\lambda + x)}{h(\lambda - x)} = \frac{\alpha}{1 - \alpha}.$$

As a consequence, the only remaining possibility is $\alpha = 1/2$. For fixed $x \in \mathbb{R}$, we have

$$1 = \frac{h(+\infty)}{h(+\infty)} = \lim_{\lambda \rightarrow \infty} \frac{h(\lambda + x)}{h(\lambda - x)} = -\frac{F_0(-x) - 1/2}{F_0(x) - 1/2}$$

implying that h must be constant using (C.5), and that F_0 must be symmetric around its median, i.e. $F_0(x) = 1 - F_0(-x)$ for all $x \in \mathbb{R}$.¹⁶ Moreover, since

¹⁶This equation implies that F_0 is necessarily continuous. This fact also follows directly from Assumption (S1) and the assumption that \mathcal{F} is closed under translations of F_0 . Indeed, assume that F_0 is discontinuous at some point x_0 . Then h has to be discontinuous at that point. But since F_0 has at most countably many points of discontinuity, there is some $\lambda_0 \in \mathbb{R}$ such that F_{λ_0} is continuous at x_0 . But this would imply that the derivative of $\bar{S}(\cdot, F_{\lambda_0})$ is discontinuous at x_0 , which contradicts the assumptions.

h is constant, (C.5) implies that also any other distribution $F \in \mathcal{F}$ must be symmetric around its median, i.e. $F(T_{1/2}(F) + x) = 1 - F(T_{1/2}(F) - x)$ for all $x \in \mathbb{R}$. However, if F_0 is symmetric around its median, then any translation F_λ of F_0 is symmetric around its median. But then, there is a convex combination of F_0 and F_λ with mixture-parameter $\beta \in (0, 1)$, $\beta \neq 1/2$, such that $\beta F_0 + (1 - \beta)F_\lambda$ is not symmetric around its median if $\lambda \neq 0$. Consequently, the conditions of the proposition are violated such that a strictly metrically \mathcal{F} -order-sensitive function for the median does not exist in this setting. \square

Proof of Proposition 3.13. Let $|x| > |z|$. Note that due to the convexity of Φ , it holds that $\Psi_x \geq \Psi_z$. Let $F \in \mathcal{F}$ with center of symmetry $c = C(F)$ and let $Y \sim F$. Then, using the fact that Φ is even and that $Y - c \stackrel{d}{=} c - Y$, one obtains

$$\begin{aligned} \bar{S}(c + x, F) - \bar{S}(c + z, F) &= \mathbb{E}_F[\Phi(x - (Y - c)) - \Phi(z - (Y - c))] \\ &= \mathbb{E}_F[\Psi_x(Y - c) - \Psi_z(Y - c)] > 0. \end{aligned}$$

This shows the strict metrical \mathcal{F} -order-sensitivity. The strict \mathcal{F} -consistency follows upon taking $z = 0$. \square

Proof of Proposition 3.15. Under the assumptions, Osband's principle yields the existence of a function $h: \text{int}(\mathbf{A}) \rightarrow \mathbb{R}$, $h > 0$ (by an argument like in the proof of Proposition 3.7) such that for all $F \in \mathcal{F}$, $x \in \text{int}(\mathbf{A})$

$$\frac{d}{dx} \bar{S}(x, F) = h(x) \bar{V}(x, F).$$

Using the same argument as in the proof of Osband's principle (Fissler and Ziegel, 2016, Theorem 3.2), h is twice differentiable. Assume that S is metrically \mathcal{F} -order sensitive. Then, due to Lemma B.3, for any $F \in \mathcal{F}$ the function $g_F: \mathbf{A} \ni x \mapsto g_F(x) = \bar{S}(T_\tau(F) + x, F)$ is an even function. Hence, invoking the smoothness assumptions, the third derivative of g_F must be odd. So necessarily $g_F'''(0) = 0$. Denoting $t_F = T_\tau(F)$, some tedious calculations lead to

$$g_F'''(0) = 2h'(t_F)(F(t_F)(1 - 2\tau) + \tau) + 2h(t_F)f(t_F)(1 - 2\tau). \quad (\text{C.7})$$

Recalling that $h > 0$ and $\tau \neq 1/2$ implies $g_{F_1}'''(0) \neq g_{F_2}'''(0)$. So S cannot be metrically \mathcal{F} -order-sensitive. \square

Remark C.1. Inspecting the proof of Proposition 3.15, equation (C.7) yields for $\tau = 1/2$

$$g_F'''(0) = h'(t_F)$$

for any $F \in \mathcal{F}$, $t_F = T_\tau(F)$. With the surjectivity of T_τ this proves that $h' = 0$, such that h is necessarily constant. Hence, we get an alternative proof that the squared loss is the only strictly metrically order-sensitive scoring function for the mean, up to equivalence.

Proof of Corollary 3.16. The linearity of T implies that T is mixture-continuous. Then the assertion follows directly by Proposition 2.4 and the special form of the image of the path γ in the proof therein, which is a line segment. \square

Proof of Proposition 3.17. Let $F \in \mathcal{F}$, $t = T(F)$, $v \in \mathbb{S}^{k-1}$ and $0 \leq s < s'$ such that $t + sv, t + s'v \in \mathbf{A}$. Then $\bar{S}(t + sv, F) = \bar{q}(F)(-\phi(t + sv) + s\nabla\phi(t + sv)v)$. The subgradient inequality yields

$$\begin{aligned} \bar{S}(t + sv, F) - \bar{S}(t + s'v, F) &< \bar{q}(F)((s' - s)\nabla\phi(t + s'v)v \\ &\quad + s\nabla\phi(t + sv)v - s'\nabla\phi(t + s'v)v) \leq 0. \quad \square \end{aligned}$$

Proof of Proposition 3.18. Let S be \mathcal{F} -order-sensitive on line segments. This implies that S is \mathcal{F} -consistent. Using the revelation principle, $S': \mathbf{A}' \times \mathbb{R} \rightarrow \mathbb{R}$,

$$S'(m_1, m_2, y) = S(m_1, m_2 - m_1^2, y) \quad (\text{C.8})$$

is an \mathcal{F} -consistent scoring function for $T' = (T_1, T_2 + T_1^2): \mathcal{F} \rightarrow \mathbf{A}'$, the pair of the first and second moment. Moreover, S' fulfils the same regularity conditions as S . Fissler and Ziegel (2016, Proposition 4.4) holds *mutatis mutandis* also for consistent scoring functions with ϕ convex. It is straight forward to check that the conditions for Fissler and Ziegel (2016, Proposition 4.4) are fulfilled for S' and T' with the canonical identification function $V': \mathbf{A}' \times \mathbb{R} \rightarrow \mathbb{R}^2$, $V'(m_1, m_2, y) = (m_1 - y, m_2 - y^2)^\top$. Hence, S' is necessarily of the form

$$S'(m_1, m_2, y) = -\phi(m_1, m_2) + \nabla\phi(m_1, m_2) \begin{pmatrix} m_1 - y \\ m_2 - y^2 \end{pmatrix} + a(y),$$

where $a: \mathbb{R} \rightarrow \mathbb{R}$ is some \mathcal{F} -integrable function and $\phi: \mathbf{A}' \rightarrow \mathbb{R}$ is a convex C^3 -function with gradient $\nabla\phi$ (considered as a row vector) and Hessian $\nabla^2\phi = (\phi_{ij})_{i,j=1,2}$. In summary, (C.8) yields the form at (3.8).

Now, we verify conditions (3.9) and (3.10). Let $F \in \mathcal{F}$, with $(t_1, t_2) = T(F)$. For $v \in \mathbb{R}^2$, $\|v\| = 1$, $s \in \mathbb{R}$ with $\bar{s} := t + sv \in \mathbf{A}$, it holds that

$$\begin{aligned} \frac{d}{ds} \bar{S}(t + sv, F) &= s(v_1, v_2 + 2v_1\bar{s}_1) \nabla^2\phi(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) \begin{pmatrix} v_1 \\ v_2 + 2v_1\bar{s}_1 - sv_1^2 \end{pmatrix} \\ &= s(v_1, v_2 + 2v_1\bar{s}_1) \nabla^2\phi(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) \begin{pmatrix} v_1 \\ v_2 + 2v_1\bar{s}_1 \end{pmatrix} \quad (\text{C.9}) \end{aligned}$$

$$- s^2 v_1^3 (\phi_{12}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) + 2\bar{s}_1 \phi_{22}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2)) \quad (\text{C.10})$$

$$- s^2 v_1^2 v_2 \phi_{22}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2). \quad (\text{C.11})$$

Since $(\phi_{ij})_{i,j=1,2}$ is positive semi-definite, the term at (C.9) is non-negative and the term at (C.11) has the sign of $-v_2$. Consider $v = (1, 0)^\top$. Due to the surjectivity of T it holds that for all $c_1 \in \mathbb{R}$, $c_2 > 0$, $s \in \mathbb{R}$ there exists a distribution $F^+ \in \mathcal{F}$ such that $T_1(F^+) + s = c_1$ and $T_2(F^+) = c_2$. Hence,

$$\begin{aligned} s^{-1} \frac{d}{ds} \bar{S}(t + sv, F^+) &= (1, 2c_1) \nabla^2\phi(c_1, c_2 + c_1^2) \begin{pmatrix} 1 \\ 2c_1 \end{pmatrix} \\ &\quad - s (\phi_{12}(c_1, c_2 + c_1^2) + 2c_1 \phi_{22}(c_1, c_2 + c_1^2)). \quad (\text{C.12}) \end{aligned}$$

Due to the \mathcal{F} -order-sensitivity of S , the term on the left-hand side of (C.12) is non-negative for all $s \in \mathbb{R}$. Since $|s|$ can be arbitrarily large, the term $\phi_{12}(c_1, c_2 + c_1^2) + 2c_1 \phi_{22}(c_1, c_2 + c_1^2)$ must vanish and we obtain (3.9).

Finally, let v be such that $v_1, v_2 \neq 0$ and w.l.o.g. $v_2 > 0$. Then

$$\begin{aligned}
& s^{-1} \frac{d}{ds} \bar{S}(\bar{s}, F) \\
&= (v_1, v_2 + 2v_1 \bar{s}_1) \nabla^2 \phi(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) \begin{pmatrix} v_1 \\ v_2 + 2v_1 \bar{s}_1 \end{pmatrix} - sv_1^2 v_2 \phi_{22}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) \\
&= v_1^2 \phi_{11}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) - sv_1^2 v_2 \phi_{22}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) \\
&\quad + (v_2 + 2v_1 \bar{s}_1) [2v_1 \phi_{12}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) + 4v_1 \bar{s}_1 \phi_{22}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) \\
&\quad + (v_2 - 2v_1 \bar{s}_1) \phi_{22}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2)] \\
&= v_1^2 \left(\phi_{11}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) - (sv_2 + 4\bar{s}_1^2) \phi_{22}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2) \right) + v_2^2 \phi_{22}(\bar{s}_1, \bar{s}_2 + \bar{s}_1^2).
\end{aligned}$$

Due to the surjectivity of T it holds that for all $c_1 \in \mathbb{R}$, $c_2 > 0$, $s < c_2/v_2$ there exists a distribution $F^+ \in \mathcal{F}$ such that $T_1(F^+) + sv_1 = c_1$ and $T_2(F^+) + sv_2 = c_2$. Consequently, one obtains the lower bound

$$\begin{aligned}
s^{-1} \frac{d}{ds} \bar{S}(\bar{s}, F^+) &\geq v_1^2 \left(\phi_{11}(c_1, c_2 + c_1^2) - (c_2 + 4c_1^2) \phi_{22}(c_1, c_2 + c_1^2) \right) \\
&\quad + v_2^2 \phi_{22}(c_1, c_2 + c_1^2)
\end{aligned}$$

and this bound is asymptotically attained for $s \uparrow c_2/v_2$. As v_2 can be arbitrarily small, it is necessary and sufficient for order sensitivity on line segments that the map $\mathbf{A} \ni (c_1, c_2) \mapsto \phi_{11}(c_1, c_2 + c_1^2) - (c_2 + 4c_1^2) \phi_{22}(c_1, c_2 + c_1^2)$ is non-negative which is equivalent to (3.10). The reverse direction follows with analogous considerations. \square

Proof of Proposition 3.20. Let $F \in \mathcal{F}$ with density f , $t = (t_1, t_2) = T(F)$, $v = (v_1, v_2) \in \mathbb{S}^2$, and $s > 0$ such that $t + sv \in \mathbf{A}$. Then, after some calculation, we find

$$\begin{aligned}
\frac{d}{ds} \bar{S}(t_1 + sv_1, t_2 + sv_2, F) &= (F(t_1 + sv_1) - \alpha) v_1 (g'(t_1 + sv_1) + \frac{1}{\alpha} \phi'(t_2 + sv_2)) \\
&\quad + sv_2^2 \phi''(t_2 + sv_2) + v_2 \phi''(t_2 + sv_2) \left(\frac{1}{\alpha} \int_{t_1}^{t_1 + sv_1} F(y) dy - sv_1 \right). \quad (\text{C.13})
\end{aligned}$$

We have

$$\alpha sv_1 \leq \int_{t_1}^{t_1 + sv_1} F(y) dy \leq F(t_1 + sv_1) sv_1. \quad (\text{C.14})$$

Note that by assumption $g' \geq 0$, $\phi' > 0$, $\phi'' > 0$, and, furthermore $(F(t_1 + sv_1) - \alpha) v_1 > 0$ for $v_1 \neq 0$. Therefore, if $v_2 \geq 0$, the first two summands on the right-hand side of (C.13) are strictly positive and the last one is non-negative using the first inequality in (C.14). For $v_2 < 0$, we find using the second inequality in

(C.14)

$$\begin{aligned}
& \frac{d}{ds} \bar{S}(t_1 + sv_1, t_2 + sv_2, F) \\
& > (F(t_1 + sv_1) - \alpha) v_1 \frac{1}{\alpha} \phi'(t_2 + sv_2) + v_2 \frac{1}{\alpha} \phi''(t_2 + sv_2) (F(t_1 + sv_1) - \alpha) sv_1 \\
& = \frac{1}{\alpha} (F(t_1 + sv_1) - \alpha) v_1 (\phi'(t_2 + sv_2) + sv_2 \phi''(t_2 + sv_2)) \geq 0,
\end{aligned}$$

where the last inequality is due to the assumption at (3.13). \square

C.3. Proofs for Section 4

Proof of Proposition 4.7. If a random variable Y has distribution F with $F \in \mathcal{F}$, we write $F - z$ for the distribution of $Y - z$ where $z \in \mathbb{R}^k$. To show the first part, consider any $F \in \mathcal{F}$ and $z \in \mathbb{R}^k$. Then

$$0 = \mathbb{E}_F[V(T(F), Y)] = \mathbb{E}_F[V(T(F) - z, Y - z)].$$

Since V is a strict \mathcal{F} -identification function for T , $T(F - z) = T(F) - z$.

For the second part, Fissler and Ziegel (2016, Theorem 3.2) implies that there exists a matrix-valued function $h: \mathbb{R}^k \rightarrow \mathbb{R}^{k \times k}$ such that

$$\nabla \bar{S}(x, F) = h(x) \bar{V}(x, F)$$

for all $x \in \mathbb{R}^k$ and for all $F \in \mathcal{F}$. We will show that h is constant. Since $\bar{S}(x, F) - \bar{S}(x', F) = \bar{S}(x - z, F - z) - \bar{S}(x' - z, F - z)$ for all $x, x', z \in \mathbb{R}^k$ and $F \in \mathcal{F}$, we obtain by taking the gradient with respect to x

$$h(x) \bar{V}(x, F) = h(x - z) \bar{V}(x - z, F - z) = h(x - z) \bar{V}(x, F), \quad (\text{C.15})$$

where the second identity is due to the linear $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariance of V . So (C.15) is equivalent to

$$\bar{V}(x, F) \in \ker(h(x - z) - h(x)).$$

Now, one can use Assumption (V1) and Fissler and Ziegel (2016, Remark 3.1), which implies that

$$\ker(h(x - z) - h(x)) = \mathbb{R}^k.$$

Since $x, z \in \mathbb{R}^k$ were arbitrary, the function h is constant. \square

Proof of Lemma 4.8. If S has linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant score differences, S satisfies (4.3) for all $x, x', y, z \in \mathbb{R}^k$. Due to Lemma 4.3, T must be $\pi_{\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k}}$ -equivariant, hence, $T(\delta_y) - z = T(\delta_{y-z})$. This yields that S_0 defined at (4.5) is linearly $(\text{id}_{\mathbb{R}^k}, \text{id}_{\mathbb{R}^k})$ -invariant. Since S and S_0 are of equivalent form, also S_0 is strictly \mathcal{F} -consistent for T . The non-negativity follows directly from the fact that \mathcal{F} contains all point measures and from the strict consistency. \square

Proof of Proposition 4.9. Fissler and Ziegel (2016, Theorem 3.2) asserts that there is a matrix-valued function $h: \text{int}(\mathbf{A}) \rightarrow \mathbb{R}^{2 \times 2}$ such that for all $(x_1, x_2) \in \text{int}(\mathbf{A})$ and for all $F \in \mathcal{F}$ we have

$$\nabla \bar{S}(x, F) = h(x_1, x_2) \bar{V}(x_1, x_2, F). \quad (\text{C.16})$$

Due to the special form of V and Assumption (F1), this equation holds also pointwise for all $y \in \mathbb{R}$. Moreover, the function h is continuously differentiable. Assume that S has linearly (M_O, M_A) -invariant score differences. This implies, combined with the previous result, that for all $(x_1, x_2) \in \text{int}(\mathbf{A})$, and for all $y, z \in \mathbb{R}$

$$\begin{aligned} h(x_1, x_2) V(x_1, x_2, y) &= \nabla_x S(x_1, x_2, y) \\ &= \nabla_x S(x_1 + z, x_2, y + z) \\ &= h(x_1 + z, x_2) V(x_1 + z, x_2, y + z) \\ &= h(x_1 + z, x_2) V(x_1, x_2, y). \end{aligned}$$

An application of Assumption (V1) (similarly to the proof of Proposition 4.7) yields that h must be necessarily constant in its first argument.

On the other hand, arguing as in the proof of Proposition 3.18 with identification function V^* , the revelation principle yields that

$$S(x_1, x_2, y) = -\phi(x_1, x_2 + x_1^2) + \nabla \phi(x_1, x_2 + x_1^2) \begin{pmatrix} x_1 - y \\ x_2 + x_1^2 - y^2 \end{pmatrix} + a(y), \quad (\text{C.17})$$

where $a: \mathbb{R} \rightarrow \mathbb{R}$ is some \mathcal{F} -integrable function and, due to our assumptions and Fissler and Ziegel (2016, Proposition 4.4), $\phi: \mathbf{A}' \rightarrow \mathbb{R}$ is C^3 and convex with gradient $\nabla \phi$ and Hessian $(\phi_{ij})_{i,j=1,2}$. Using the representation at (C.17), one obtains $\partial_2 S(x_1, x_2, y) = \phi_{22}(x_1, x_2 + x_1^2)(x_2 + x_1^2 - y^2)$. A comparison to the form at (C.16) yields that

$$\begin{aligned} h_{22}(x_1, x_2) &= \phi_{22}(x_1, x_2 + x_1^2) \\ h_{21}(x_1, x_2) &= 2x_1 \phi_{22}(x_1, x_2 + x_1^2). \end{aligned}$$

Since $\partial_1 h_{22}(x_1, x_2)$ vanishes, we obtain that $0 = \partial_1 h_{21}(x_1, x_2) = 2\phi_{22}(x_1, x_2 + x_1^2)$. As the Hessian of ϕ must be positive semi-definite $\phi_{11} \geq 0$ and $\phi_{12} = \phi_{21} = 0$. Since ϕ is C^3 , we have that $\partial_2 \phi_{11} = \partial_1 \phi_{12} = 0$, hence ϕ_{11} is constant in the first argument. Equating the first component of (C.16) and (C.17) and using that $h_{12} = h_{21} = 0$, we find that $\phi_{11}(x_1, x_2 + x_1^2) = h_{11}(x_1, x_2)$. As h_{11} is constant in x_1 , this implies that ϕ_{11} is also constant in its second argument which yields the claim. \square

Proof of Proposition 4.10. The scoring function S_c is of equivalent form as given at (3.12) with $g(x_1) = -x_1^2/2 + cx_1$ and $\phi(x) = (\alpha/2)x_2^2$. This means that ϕ is strictly convex and the function $x_1 \mapsto x_1 \phi'(x_2)/\alpha + g(x_1)$ is strictly increasing in x_1 if and only if $x_2 + c > x_1$, that is, if and only if $(x_1, x_2) \in \mathbf{A}_c$. Moreover, one can verify that the action domain \mathbf{A}_c satisfies the conditions introduced in

Fissler and Ziegel (2019, Proposition 2). Hence, the latter proposition implies the strict \mathcal{F} -consistency of S_c . Finally, a direct computation yields that $S_c(x_1 + z, x_2 + z, y + z) = S_c(x_1, x_2, y)$ for all $(x_1, x_2) \in \mathbf{A}_c$, $y, z \in \mathbb{R}$. This proves the first part.

Under the conditions of Fissler and Ziegel (2016, Theorem 5.2(iii)), any strictly \mathcal{F} -consistent scoring function $S: \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$, where $\mathbf{A} = T(\mathcal{F})$, is almost everywhere of the form given at (3.12) with g continuously differentiable and ϕ twice continuously differentiable. By translation invariance of score differences the function $\Psi: \mathbb{R} \times \mathbf{A} \times \mathbf{A} \times \mathbb{R} \rightarrow \mathbb{R}$,

$$\begin{aligned} \Psi(z, x_1, x_2, x'_1, x'_2, y) &= S(x_1 + z, x_2 + z, y + z) - S(x'_1 + z, x'_2 + z, y + z) \\ &\quad - S(x_1, x_2, y) + S(x'_1, x'_2, y) \end{aligned}$$

constantly vanishes. Let $z, y \in \mathbb{R}$ and $(x_1, x_2), (x'_1, x'_2) \in \mathbf{A}$. Then

$$\begin{aligned} 0 &= \frac{d}{dx_2} \Psi(z, x_1, x_2, x'_1, x'_2, y) \\ &= (x_2 - x_1 + \frac{1}{\alpha} \mathbb{1}\{y \leq x_1\}(x_1 - y))(\phi''(x_2 + z) - \phi''(x_2)), \end{aligned}$$

hence ϕ'' is constant, that is, $\phi(x_2) = d_1 x_2^2 + d_2 x_2 + d_3$ with $d_1 > 0$ (ensuring the strict convexity of ϕ) and $d_2, d_3 \in \mathbb{R}$. Similarly, the derivative of Ψ with respect to z must vanish for all $z, y \in \mathbb{R}$ and $(x_1, x_2), (x'_1, x'_2) \in \mathbf{A}$. A calculation yields

$$\begin{aligned} 0 &= \frac{d}{dz} \Psi(z, x_1, x_2, x'_1, x'_2, y) \\ &= (\mathbb{1}\{y \leq x_1\} - \alpha)g'(x_1 + z) - \mathbb{1}\{y \leq x_1\}g'(y + z) \\ &\quad - (\mathbb{1}\{y \leq x'_1\} - \alpha)g'(x'_1 + z) + \mathbb{1}\{y \leq x'_1\}g'(y + z) \\ &\quad + \frac{2d_1}{\alpha}(\mathbb{1}\{y \leq x_1\}(x_1 - y) - x_1) - \frac{2d_1}{\alpha}(\mathbb{1}\{y \leq x'_1\}(x'_1 - y) - x'_1). \end{aligned}$$

This implies that necessarily $g'(x_1) = (-2d_1/\alpha)x_1 + d_4$ for some $d_4 \in \mathbb{R}$. Hence, $g(x_1) = (-d_1/\alpha)x_1^2 + d_4x_1 + d_5$ for some $d_5 \in \mathbb{R}$. Now, by Fissler and Ziegel (2016, Theorem 5.3(iii)), the function

$$\psi_{x_2}(x_1) = x_1\phi'(x_2)/\alpha + g(x_1) = x_1(2d_1x_2 + d_2)/\alpha - d_1x_1^2/\alpha + d_4x_1 + d_5$$

must be strictly increasing in x_1 which holds if and only if

$$x_2 + \frac{d_2 + d_4\alpha}{2d_1} > x_1.$$

This condition is satisfied for all $(x_1, x_2) \in \mathbf{A} = T(\mathcal{F})$ if and only there is a $c > 0$ such that $T(\mathcal{F}) \subseteq \mathbf{A}_c$ and $d := (d_2 + d_4\alpha)/(2d_1) \geq c$. The scoring function at (3.12) with $\phi(x_2) = d_1x_2^2 + d_2x_2 + d_3$, $d_1 > 0$, $d_2, d_3 \in \mathbb{R}$, $g(x_1) = (-d_1/\alpha)x_1^2 + d_4x_1 + d_5$, $d_4, d_5 \in \mathbb{R}$ is equivalent to S_d defined at (4.7), which concludes the proof. \square

Proof of Proposition 4.11. Suppose ϕ satisfies (4.12). This implies that for any $c > 0$ the map $z \mapsto \phi(\Lambda(c)z) - c^b\phi(z)$ is an affine function. Moreover, a Taylor expansion yields that for all $x, z \in \mathbf{A}$

$$\phi(\Lambda(c)z) - c^b\phi(z) = (\nabla\phi(\Lambda(c)x)\Lambda(c) - c^b\nabla\phi(x))(z - x) + \phi(\Lambda(c)x) - c^b\phi(x).$$

Then, a direct calculation yields the result.

Now, suppose (4.10) is satisfied. Its left-hand side equals

$$\begin{aligned} & -\phi(\Lambda(c)x) + \nabla\phi(\Lambda(c)x)\Lambda(c)x + \phi(\Lambda(c)z) - \nabla\phi(\Lambda(c)z)\Lambda(c)z \\ & + \left(\nabla\phi(\Lambda(c)z)\Lambda(c) - \nabla\phi(\Lambda(c)x)\Lambda(c) \right) (y, y^2, \dots, y^k)^\top, \end{aligned}$$

whereas the right-hand side is

$$\begin{aligned} & -c^b\phi(x) + c^b\nabla\phi(x)x + c^b\phi(z) - c^b\nabla\phi(z)z \\ & + c^b \left(\nabla\phi(z) - \nabla\phi(x) \right) (y, y^2, \dots, y^k)^\top. \end{aligned}$$

Both terms are polynomials in y of degree k , which leads to the identity

$$\nabla\phi(\Lambda(c)z)\Lambda(c) - \nabla\phi(\Lambda(c)x)\Lambda(c) = c^b \left(\nabla\phi(z) - \nabla\phi(x) \right).$$

This is exactly condition (4.12). \square

Proof of Corollary 4.12. The form at (4.13) follows as in the proof of Proposition 3.18. The rest follows by Proposition 4.11. \square

Acknowledgments

We would like to thank Ingo Steinwart, Tilmann Gneiting, Werner Ehm, Sebastian Lerch, Alexander Jordan, Fabian Krüger, and Jonas Brehmer for stimulating and insightful discussions about the topic.

We are grateful to the editor, an anonymous associate editor and an anonymous referee for their valuable comments which helped to improve the presentation of the paper.

References

- ABERNETHY, J. D. and FRONGILLO, R. (2012). A Characterization of Scoring Rules for Linear Properties. In *Proceedings of the 25th Annual Conference on Learning Theory. Proceedings of Machine Learning Research* **23** 27.1–27.13. PMLR, Edinburgh, Scotland.
- ACERBI, C. and SZEKELY, B. (2014). Backtesting Expected Shortfall. *Risk Magazine* **27** 76–81.
- ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. Springer, Berlin Heidelberg New York. [MR2378491](#)

- BANERJEE, A., GUO, X. and WANG, H. (2005). On the Optimality of Conditional Expectation as a Bregman Predictor. *IEEE Transactions on Information Theory* **51** 2664–2669. [MR2246384](#)
- BELLINI, F. and BIGNOZZI, V. (2015). On elicitable risk measures. *Quantitative Finance* **15** 725–733. [MR3334566](#)
- BREHMER, J. R. (2017). Elicitability and its Application in Risk Management, Master's thesis, University of Mannheim.
- DIEBOLD, F. X. and MARIANO, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13** 253–263. [MR3303732](#)
- EHM, W., GNEITING, T., JORDAN, A. and KRÜGER, F. (2016). Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78** 505–562. [MR3506792](#)
- EMBRECHTS, P. and HOFERT, M. (2014). Statistics and Quantitative Risk Management for Banking and Insurance. *Annual Review of Statistics and Its Application* **1** 493–514.
- EMBRECHTS, P., PUCCETTI, G., RÜSCHENDORF, L., WANG, R. and BELERAJ, A. (2014). An Academic Response to Basel 3.5. *Risks* **2** 25–48.
- FISSLER, T. (2017). On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space, PhD thesis, University of Bern.
- FISSLER, T. and ZIEGEL, J. F. (2016). Higher order elicibility and Osband's principle. *Annals of Statistics* **44** 1680–1707. [MR3519937](#)
- FISSLER, T. and ZIEGEL, J. F. (2019). Erratum: Higher Order Elicitability and Osband's Principle. *arXiv* **1901.08826v1**. [MR3519937](#)
- FRIEDMAN, D. (1983). Effective Scoring Rules for Probabilistic Forecasts. *Management Science* **29** 447–454.
- FRONGILLO, R. and KASH, I. (2015a). Vector-Valued Property Elicitation. In *Proceedings of The 28th Conference on Learning Theory. Proceedings of Machine Learning Research* **40** 710–727. PMLR, Paris, France.
- FRONGILLO, R. and KASH, I. (2015b). On Elicitation Complexity. In *Advances in Neural Information Processing Systems 28* (C. CORTES, N. D. LAWRENCE, D. D. LEE, M. SUGIYAMA and R. GARNETT, eds.) 3258–3266. Curran Associates, Inc.
- GNEITING, T. (2011). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association* **106** 746–762. [MR2847988](#)
- HEINRICH, C. (2014). The mode functional is not elicitable. *Biometrika* **101** 245–251. [MR3180670](#)
- HUBER, P. J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics* 73–101. [MR0161415](#)
- HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey. [MR2488795](#)
- KOENKER, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge. [MR2268657](#)
- KÖNIGSBERGER, K. (2004). *Analysis 2*, 5th ed. Springer-Verlag, Berlin Heidelberg New York. [MR1251736](#)
- LAMBERT, N. (2013). Elicitation and Evaluation of Statistical Functionals.

Preprint.

- LAMBERT, N., PENNOCK, D. M. and SHOHAM, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce* 129–138. ACM, Chicago, IL, USA.
- LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer Science & Business Media, New York. [MR1639875](#)
- MURPHY, A. H. and DAAN, H. (1985). Forecast Evaluation. In *Probability, Statistics and Decision Making in the Atmospheric Sciences* (A. H. MURPHY and R. W. KATZ, eds.) 379–437. Westview Press, Boulder, Colorado.
- NAU, R. F. (1985). Should Scoring Rules Be ‘Effective’? *Management Science* **31** 527–535.
- NEWAY, W. K. and POWELL, J. L. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica* **55** 819–847. [MR0906565](#)
- NOLDE, N. and ZIEGEL, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics*. To appear. [MR3743276](#)
- OSBAND, K. H. (1985). Providing Incentives for Better Cost Forecasting, PhD thesis, University of California, Berkeley.
- PATTON, A. J. (2011). Data-based ranking of realised volatility estimators. *Journal of Econometrics* **161** 284–303. [MR2774942](#) [MR2774942](#)
- PATTON, A. J. (2017). Comparing Possibly Misspecified Forecasts. Working paper, Duke University.
- REICHELSTEIN, S. and OSBAND, K. (1984). Incentives in government contracts. *Journal of Public Economics* **24** 257–270.
- SAERENS, M. (2000). Building cost functions minimizing to some summary statistics. *IEEE Transactions on Neural Networks* **11** 1263–1271.
- SAVAGE, L. J. (1971). Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association* **66** 783–801. [MR0331571](#)
- SCOTT, J. and MARSHALL, G. (2009). *A Dictionary of Sociology*, 3 rev. ed. Oxford University Press.
- STEINWART, I. (2007). How to Compare Different Loss Functions and Their Risks. *Constructive Approximation* **26** 225–287. [MR2327600](#)
- STEINWART, I., PASIN, C., WILLIAMSON, R. and ZHANG, S. (2014). Elicitation and Identification of Properties. In *Proceedings of The 27th Conference on Learning Theory. Proceedings of Machine Learning Research* **35** 482–526. PMLR, Barcelona, Spain.
- THOMSON, W. (1979). Eliciting production possibilities from a well-informed manager. *Journal of Economic Theory* **20** 360–380. [MR0540822](#)
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. [MR1652247](#)
- WEBER, S. (2006). Distribution-Invariant Risk Measures, Information, and Dynamic Consistency. *Mathematical Finance* **16** 419–441. [MR2212272](#)